A UNITED STATES
DEPARTMENT OF
COMMERCE
PUBLICATION

NOAA Technical Memorandum ERL OD-15

U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
Environmental Research Laboratories

On the Use of Gamma Functions and Bayesian Analysis in Evaluating Florida Cumulus Seeding Results

JOANNE SIMPSON
JANE C. EDEN
ANTHONY OLSEN
JACQUES PÉZIER

Office
of the Director
BOULDER,
COLORADO
February 1973

ENVIRONMENTAL RESEARCH LABORATORIES

OFFICE OF THE DIRECTOR

IMPORTANT NOTICE

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

BOULDER, COLORADO

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
Environmental Research Laboratories

NOAA Technical Memorandum ERL OD-15

ON THE USE OF GAMMA FUNCTIONS AND BAYESIAN ANALYSIS IN EVALUATING FLORIDA CUMULUS SEEDING RESULTS

Joanne Simpson
Jane C. Eden
Anthony Olsen

Experimental Meteorology Laboratory

Jacques Pézier

C.E.S.A., Jouy en Josas, France
Consultant to ERL-NOAA

Office of the Director
Boulder, Colorado
February 1973

TABLE OF CONTENTS

ABSTRACT

Bayesian techniques are used to evaluate the seeding factor on rainfall and its probability distribution in the Experimental Meteorology Laboratory randomized dynamic seeding experiments on isolated cumuli in Florida. A framework is constructed for later use of these tools with the randomized multiple cumulus seeding experiment in the 4000 n mi^2 target area.

The data used are the cloud base integrated rainfall volumes obtained by a calibrated 10-cm radar which was tested against a gage network. The analyses are based on the finding that the seeded and control data are both well fitted by a gamma distribution with the same shape parameter. Since seeding alters only the scale parameter, Bayes equation is adapted to finding the distribution of the scale parameter and/or the seeding factor directly. The natural distribution is determined by the best-fit gamma function to the control cloud data. A wide variety of prior distributions of scale parameter are used, with prior expected seeding factors ranging from 0.5 to three. After seeing the seeded cloud data, posterior expected seeding factors range from slightly below two to slightly above three, depending upon the prior distribution assumed. The 95 percent integrated probability range for seeding factor is established, which is not extended to values less than one (reduction in rainfall) even under absurdly biased initial assumptions.

The final section endeavors to determine how many observations are required to establish the parameters of the gamma distribution adequately for seeding factor determinations. A program is devised to generate random "rain" observations from a predetermined gamma distribution. Then the parameters are recovered in the same way as done for the data, namely with a program based on the principle of maximum entropy.

ON THE USE OF GAMMA FUNCTIONS AND BAYESIAN ANALYSIS IN EVALUATING FLORIDA CUMULUS SEEDING RESULTS

Joanne Simpson, Jane C. Eden, Anthony Olsen and Jacques Pézier

1. INTRODUCTION

The randomized single cumulus experiments in south Florida in 1968 and 1970 were only the first phase in a continued program to develop dynamic seeding as a tool in water management, in convective cloud research, and in eventual severe storm modification. The Experimental Meteorology Laboratory (EML) is now at work on an extended series of multiple cumulus seeding experiments in a 4000 n mi^2 target area, which present much greater obstacles both scientifically and operationally.

Aspects of the EML randomized dynamic seeding experiments on single cumuli have already been reported in the literature (Simpson et al., 1970; 1971; Simpson and Woodley, 1971). Classical statistical analyses showed that the seeded single clouds grew higher and rained more than the controls, with an average radar-measured rainfall difference of 270 acre-ft. Differences were significant at the 5 percent level or better, using several types of statistical tests. Most of the tests were executed on a set of "transformed" data, which were obtained by taking the fourth root of the total rainfall obtained by radar integration from each cloud. This transformation was made to eliminate the effects of extremes and to render the data distribution more nearly normal. Using this procedure, we were able to obtain a quantitative estimate of the seeding factor on single cloud rainfall, namely that it slightly exceeded three (Simpson et al., loc. cit., 1971). A particularly interesting result with the transformed data sets was that they were well fitted by gamma distributions, with virtually unchanged shape parameters between seeded and control sets (Simpson, 1972). The data are presented in table 1. Details of how they were obtained have been explained by Woodley (1970), and Woodley and Herndon (1970).

Table 1. Single Cloud Data for 1968 and 1970.

Total cloud lifetime			
Seeded rain		Control rain	
Acre-feet (raw)	Fourth root (transformed)	Acre-feet (raw)	Fourth root (transformed)
129.6	3.37405	26.1	2.26027
31.4	2.36719	26.3	2.26459
2,745.6	7.23868	87.0	3.05408
489.1	4.70272	95.0	3.12199
430.0	4.55373	372.4	4.39291
302.8	4.17147	0	0 (1)*
119.0	3.30283	17.3	2.03944
4.1	1.42297	24.4	2.22253
92.4	3.1004	11.5	1.84151
17.5	2.04531	321.2	4.23344
200.7	3.76389	68.5	2.87689
274.7	4.07113	81.2	3.00185
274.7	4.07113	47.3	2.6225
7.7	1.6658	28.6	2.31255
1,656.0	6.37918	830.1	5.36763
978.0	5.59223	345.5	4.31134
198.6	3.754	1,202.6	5.88885
703.4	5.14992	36.6	2.45963
1,697.8	6.41906	4.9	1.48782
334.1	4.27532	4.9	1.48782
118.3	3.29797	41.1	2.53198
255.0	3.99609	29.0	2.3206
115.3	3.27686	163.0	3.57311
242.5	3.94619	244.3	3.95349
32.7	2.39132	147.8	3.48673
40.6	2.52424	21.7	2.15832

* Since some of our computer programs involve products of the data and will not accept zero, one acre-ft has usually been substituted for zero for this observation.

In 1971 the senior author attended a Decision Analysis course at Dartmouth College, taught by the fourth author, with a primary purpose of adapting Bayesian analysis to the design and analysis of the EML modification experiments. Used together with the properties of gamma functions, Bayesian techniques have produced some new results for the single cloud experiments and opened up promising approaches to the multiple cloud seeding or area experiments. Here we present mainly the results from applying the techniques to the single cloud data while building up the framework to use with the area data, for which preliminary results will be presented in a sequel paper.

With the single clouds, the sample of cases is more extensive (26 seeded and 26 control cases) than we will probably be able to obtain for some years in the area experiment. Furthermore, the radar calibration for 1968 and 1970 was found (Woodley and Herndon, 1970; Herndon et al., 1971) to be quite accurate[1]. Hence these data appear ideal to adaptation and proving out the application of Bayesian techniques to cumulus modification. Errors in the data are considered in other EML reports; in this one they are assumed correct as given in table 1.

2. BACKGROUND FOR THE BAYESIAN APPROACH AND THE APPLICATION OF GAMMA DISTRIBUTIONS TO THE DATA

The Bayesian approach to seeding problems offers major attractions and also poses some difficulties. One attraction is that the use of Bayes equation permits a numerical assessment of the magnitude of a seeding factor, together with a probability distribution for it, whose standard deviation goes down as the data sample increases. This type of result contains considerably more information than does the mere rejection of the

[1]Unfortunately raingage comparisons suggest that it deteriorated in 1971, the second year of the area experiment, further complicating evaluation of that already too sparse data sample.

null hypothesis which is all that many classical tests seek to establish.

We reject the common criticism of the Bayesian approach which contends that subjective prior probabilities may introduce bias: firstly, because we usually use diffuse prior probabilities and secondly, because we test the sensitivity of our results to a very large number and many different types of prior probability distributions.

A very real difficulty, however, in applying Bayesian statistics to weather modification is that to do so we must know the distribution of natural property or properties to be modified. This requirement entails that we know the distribution function and its sufficient statistics, which generally implies two moments of the distribution. It also implies that we assume that the natural distribution function remains stationary in time. In our work, this assumption is far less dangerous than in many other uses of the approach, since we screen out all but fair convective days in south Florida summers.

A key first result was that the single cloud transformed (fourth root) rainfall data was well fitted by a gamma distribution. The gamma probability density function may be written

$$p(R) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} R^{\alpha-1} e^{-\beta R} \qquad (1)$$

where $p(R)$ is the probability density of a rainfall amount, R. The scale of the distribution is determined by the parameter β and the shape by the parameter α. Γ is the gamma function (cf. Pearson et al., 1957). The first two moments of the gamma function are well known to be (Kendall and Stewart, 1963):

$$\mu_1 = \langle R \rangle = \alpha/\beta$$

and (2)

$$\mu_2 = \sigma^2 = \alpha/\beta^2$$

where $\langle R \rangle$ is the expected value and σ^2 is the variance. Therefore, the coefficient of variation V is

$$V \equiv \frac{\sigma}{\langle R \rangle} = \frac{1}{\sqrt{\alpha}} \tag{3}$$

We developed a computer program (DAMAX) utilizing the principle of maximum entropy (Tribus, 1969, p. 197) to find the best fit distribution (from a desired number of possible distributions) for any set of data. The program also calculates the key parameters for each distribution, e.g. α and β for the gamma distribution[2] and also a chi-square measure of goodness of fit. An early version of the program is listed in a report by Simpson & Pezier (1971) and the latest version can be obtained from EML on request.

Using this program we found (Simpson, 1972) that the gamma distribution excellently fit the EML single cloud transformed rainfall data, for both seeded and control populations separately. It turned out that the coefficient of variation was virtually identical for the two populations, which differed in their expected values. Thus the shape parameter appeared to be unaffected by seeding, which only diminished the scale parameter, so that the whole distribution was moved toward higher rainfalls. With $V \simeq 0.377$, it also turns out conveniently that the shape parameter is

[2] which are shown to be identical to those obtained by classical statistical methods.

about 7, an integer, which greatly simplifies calculations using (1). Sequel work has treated the data without any transformations, hereafter called "raw" data, in which α is not an integer.

3. COMPOSITE HYPOTHESIS TESTING WITH BAYES EQUATION TO ESTIMATE THE MAGNITUDE OF THE SEEDING FACTOR - SINGLE CLOUD TRANSFORMED DATA

One of the fine features of Bayesian statistics is that it can be used to estimate the magnitude of the seeding factor and also the integrated probability that its magnitude lies between any predetermined limits. In the following we will apply Bayes equation to the scale parameter β, and directly to the seeding factor, using both raw and transformed data. Transformed or fourth root data will be denoted by primed quantities throughout, while unprimed quantities refer to raw data.

First, we use Bayes equation to obtain a probability distribution for the gamma function scale parameter β', using transformed data, namely

$$p(\beta'|D') = p(\beta') \quad p(D'|\beta')/p(D') \tag{4}$$

where $p(\beta'|D')$ is the probability density distribution of the parameter given the seeded data. $p(\beta')$ is the prior probability assignment of β'. $p(D'|\beta')$ is the probability of the data, given β', while the denominator, $p(D')$, is the probability of the data, a normalizing factor only. It is assumed that both seeded and control distributions are gamma distributions with $\alpha' = 7$ in the transformed sets, while β' (control) was evaluated as 2.38620 from $\langle R' \rangle$ control = 2.93353 (acre-ft)$^{.25}$

The next preliminary step is to relate β' to seeding factor F; results are presented in table 2 and figure 1. Note that the tabulated seeding factor always relates to raw data. We obtain the β' in the table as a function of seeding factor as follows:

$$F = \frac{\langle R \rangle \text{ seeded}}{\langle R \rangle \text{ control}} \qquad \text{R is raw rainfall in acre-ft} \tag{5}$$

$$\beta' \text{ (control)}: = \frac{\alpha'}{\langle R' \rangle}_{\text{control}} = \frac{7}{\langle R' \rangle}_{\text{control}} \tag{6}$$

$$\beta' \text{ (seeded)} = \frac{\alpha'}{\langle R' \rangle}_{\text{seeded}} \quad \text{where we assume for the moment that} \tag{7}$$

$$\langle R' \rangle \text{ seeded} = \langle F \rangle^{.25} \quad \langle R' \rangle_{\text{control}} \tag{8}$$

This approximation is discussed at the end of the current section and is shown to cause only a slight positive bias for seeding factors in the range of interest.

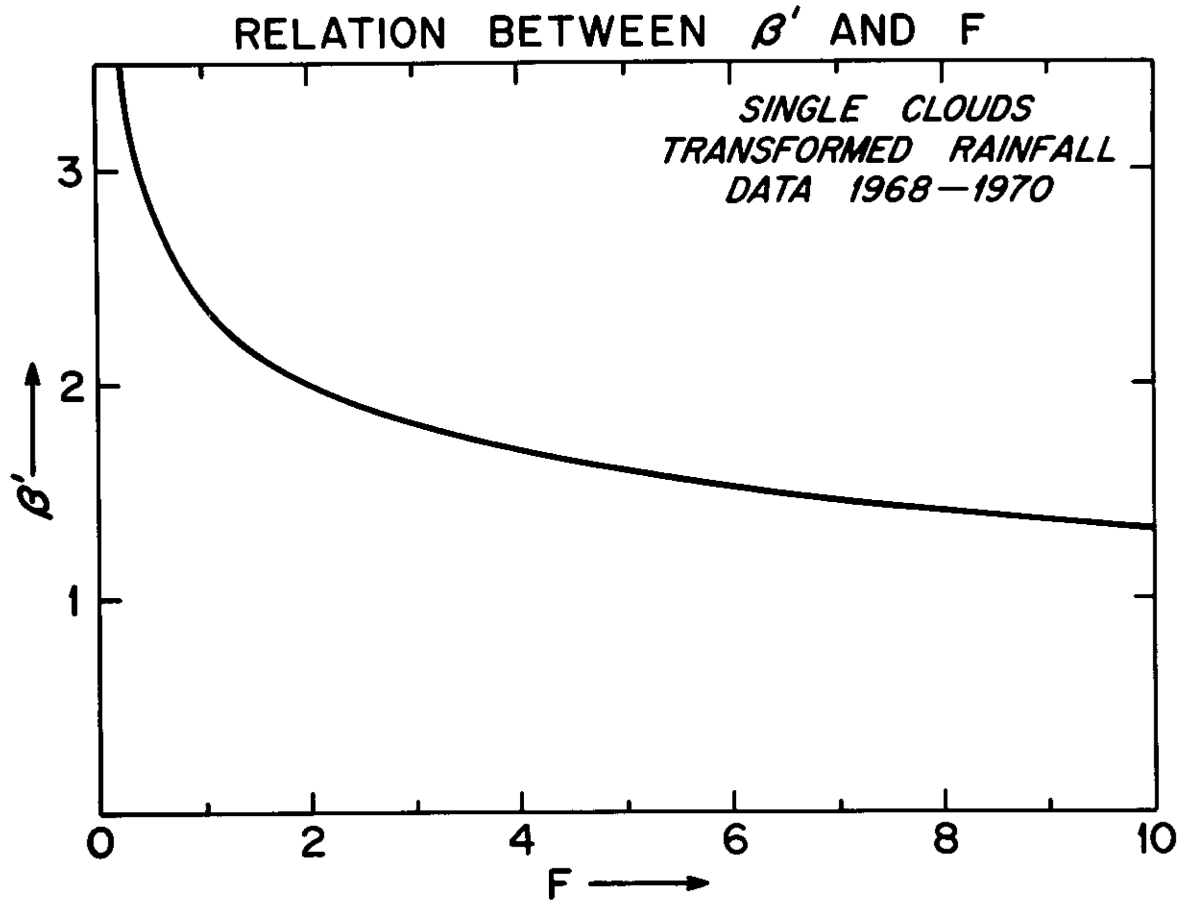

Figure 1. Graph showing relationship between seeding factor F and gamma function scale parameter β' for single cloud transformed (fourth root) data. F and β' are determined as specified by equations (5) and (6).

Table 2. Relation Between Seeding Factor F and β' Gamma Distribution's Scale Parameter.

Single Cloud Transformed Data	
$\alpha' = 7$	
F	β'
0.25	3.37460
0.5	2.83769
0.8	2.52310
1	2.38620
2	2.00650
3	1.81312
4	1.68730
5	1.59575
7	1.46701
10	1.34186

The analysis with (4) is carried out in two ways on the transformed data. The first way assumes a gamma function for the prior probability distribution of β' and the second assumes a uniform distribution for the prior.

In the first approach

$$p(\beta') = \frac{K_2^{K_1}}{\Gamma(K_1)} \beta'^{K_1 - 1} e^{-K_2\beta'} \tag{9}$$

where K_1 is the shape parameter and K_2 the scale parameter of the prior probability distribution assumed for β'.

Now we know that

$$p(D'|\beta') = \prod_{i=1}^{n} \frac{\beta'^{\alpha'}}{\Gamma(\alpha')} R_i'^{\alpha' - 1} e^{-\beta' R'_i} \tag{10}$$

where n is the number of seeded cases and R'_i is the transformed rainfall in the i^{th} seeded case. Substituting (10) and (9) into (4) we find

$$p(\beta'|D') = \frac{\left(\sum_{i=1}^{n} R'_i + K_2\right)^{n\alpha' + K_1}}{\Gamma\,(n\alpha' + K_1)} \beta'^{\,n\alpha' + K_1 - 1} e^{-\beta'\left(\sum_{i=1}^{n} R'_i + K_2\right)} \tag{11}$$

The normalizing constant is found from the exponents, since we know the resulting distribution is also a gamma distribution. Now since $n = 26$ and $\alpha' = 7$, the shape parameter of the posterior β' distribution is very large, namely here $182 + K_1$. According to Thom[3] when the scale parameter exceeds about ten, the gamma distribution degenerates, for all practical purposes, into a Gaussian distribution. Therefore, with this approach our posterior β' distributions can be treated as Gaussian, which means that the probability is about 95% that the value of β' lies within two standard deviations of the expected value. This information is used later in tables 3 and 4. We further know that

$$\langle\beta'|D'\rangle = \frac{n\alpha' + K_1}{\sum\limits_{i=1}^{n} R'_i + K_2} = \frac{\alpha' + \frac{K_1}{n}}{\overline{R'_i} + \frac{K_2}{n}} \tag{12}$$

so that

$$\lim_{n \to \text{large}} \langle\beta'|D'\rangle = \frac{\alpha'}{\overline{R'_i}} \tag{13}$$

and

$$V^2\;(\beta'|D') = \frac{1}{n\alpha' + K_1} \tag{14}$$

---

[3] Personal communication

With any gamma function for the prior probability distribution on β' we find that in the limit of large n

$$\beta'_{n \to \text{large}} = 1.80459 \text{ so that } F = 3.07 \tag{15}$$

in fine agreement with the results of the classical statistical analysis of Simpson et al., (1971). For this analysis we use two prior β' distributions, the first with $K_1 = 12$ (peaked, see fig. 2) and the second with $K_1 = 2$ (much flatter, see fig. 3).

Results for both priors with a wide range in the prior expected value of β' are presented in tables 3 and 4. Figures 2 and 3 compare plots of both prior (dashed) and posterior (solid) probability density distributions for selected cases from the tables.

Table 3. Transformed Single Cloud Data.

$K_1 = 12$ Prior β' Peaked Gamma Function			Posterior β' Gaussian	
Prior $<\beta'>$	F	Post $<\beta'>$	F	F for 95% Prob.
2.83769	0.5	1.84617	2.8	1.4 - 5.2
2.38620	1	1.83222	2.9	1.5 - 5.3
2.0065	2	1.81590	2.99	1.53 - 5.4
1.81312	3	1.80512	3.05	1.58 - 5.57
1.59575	5	1.79010	3.15	1.60 - 5.8
1.34348	10	1.76709	3.40	1.87 - 6.0

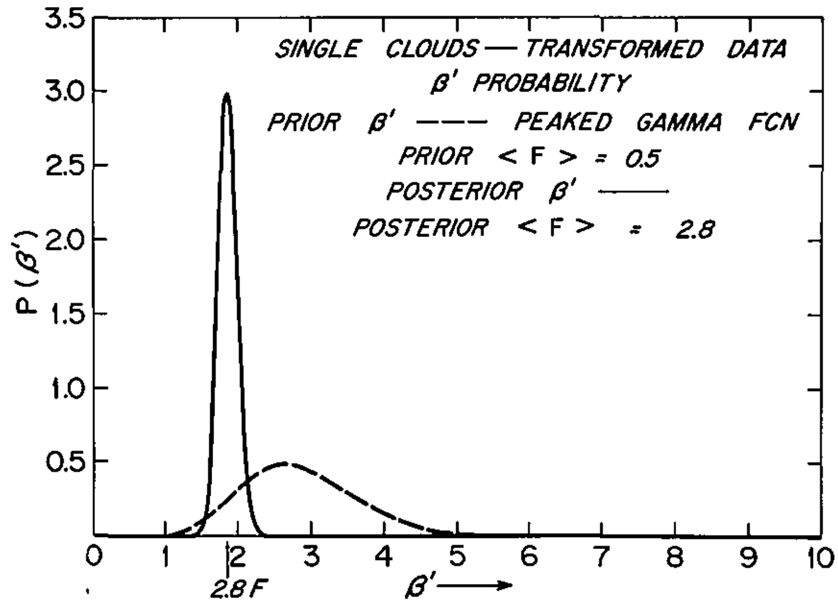

Figure 2a. Prior and posterior probability distributions of β', the scale parameter of gamma distribution. The prior probability distribution (dashed) was $K_1 = 12$ and K_2 is chosen so that the expected value of prior β corresponds to a seeding factor of 0.5 (rainfall reduction by a factor of 2). The posterior probability distribution of β' (solid) is determined by inserting the seeding rainfall data in Bayes' equation.

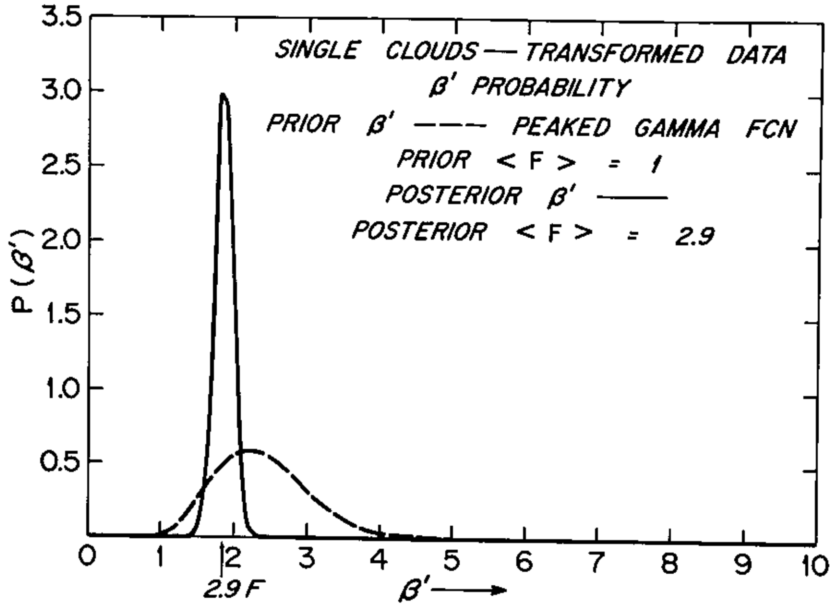

Figure 2b. Same as 2a, except that the expected value of the β' corresponds to F = 1, i.e., seeding has no effect.

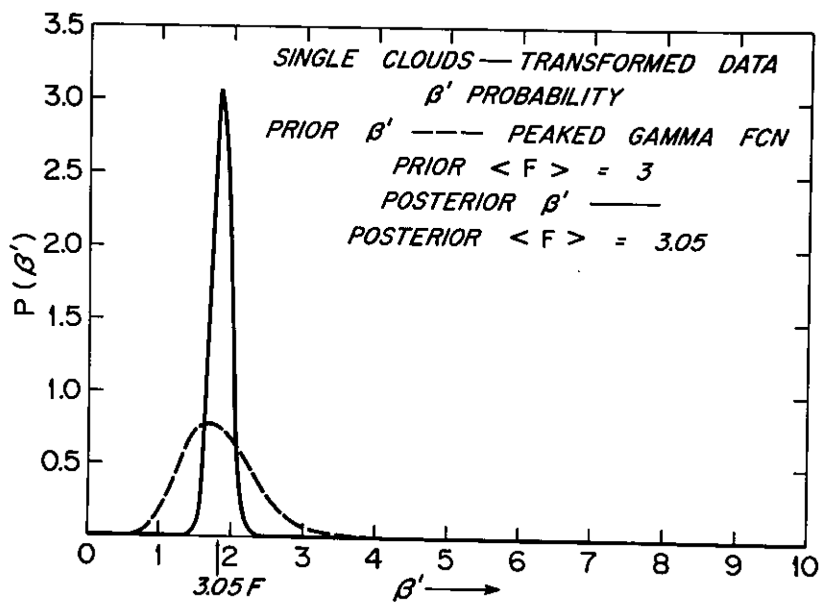

Figure 2c. Same as 2a, except that the expected value of the prior β' corresponds to F = 3, i.e., seeding multiplies the rainfall by 3.

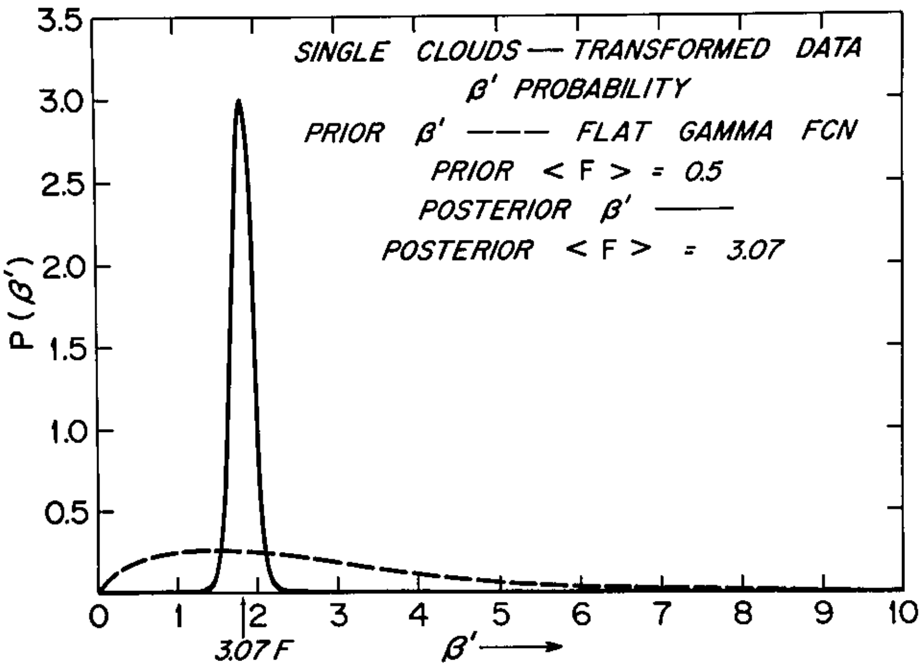

Figure 3a. *Same as figure 2 except that* $K_1 = 2$, *giving a flatter prior probability distribution of* β'.

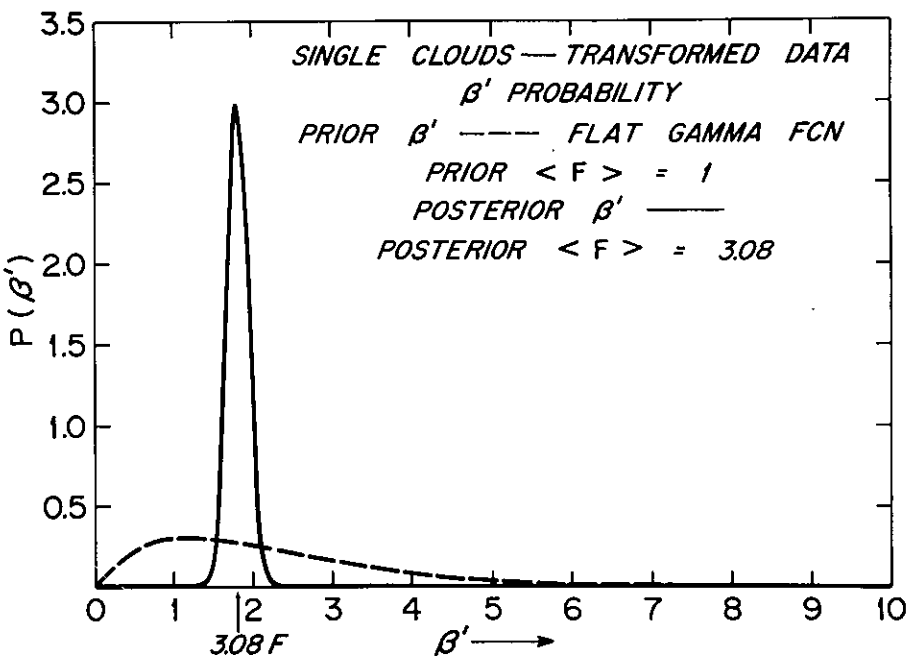

Figure 3b. *Same as figure 2 except that* $K_1 = 2$, *giving a flatter prior probability distribution of* β'.

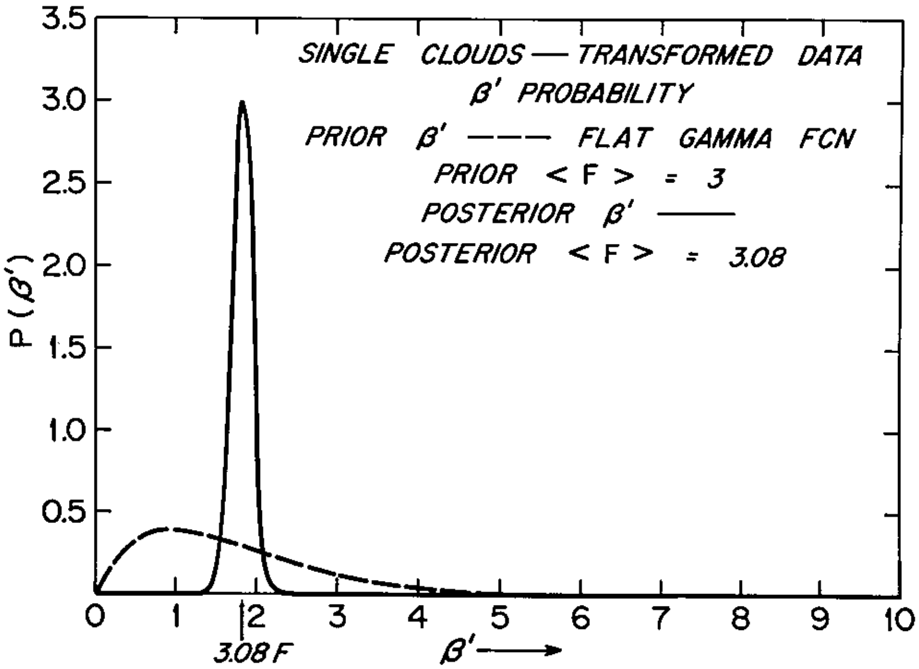

Figure 3c. Same as figure 2 except that $K_1 = 2$, giving a flatter prior probability distribution of β'.

Table 4. Transformed Single Cloud Data.

$K_1 = 2$ Prior β' Flat Gamma Function			Posterior β' Gaussian	
Prior $<\beta'>$	F	Post $<\beta'>$	F	F for 95% Prob.
2.83769	0.5	1.81176	3.07	1.65 - 5.4
2.38620	1	1.80950	3.08	1.65 - 5.65
2.0065	2	1.80657	3.08	1.65 - 5.65
1.81312	3	1.80468	3.08	1.65 - 5.65
1.59575	5	1.80203	3.08	1.65 - 5.65
1.34348	10	1.79788	3.10	1.73 - 5.80

As might be expected, when the prior β' distribution is more peaked ($K_1 = 12$) the posterior $<\beta'>$ is fairly sensitive to the prior $<\beta'>$, but when the prior β' distribution is flatter, there is virtually no sensitivity to the prior scale parameter and the seeding factor is almost identical to that obtained from classical statistics. Although the 95 percent probability permits a rather wide range in seeding factor, it is substantially positive at even the lowest limit. These results are further confirmed when we take the most diffuse prior distribution of β', namely uniform.

The most diffuse prior probability to place on β' is one which is uniform over a wide range. We start with Bayes equation in form (4) where again the denominator is regarded as a normalizing constant. We obtain:

$$p(\beta'|D') = \frac{p(D'|\beta')\ p(\beta')}{\text{Denom.}} = \frac{\dfrac{\beta'^{n\alpha'}}{[\Gamma(\alpha')]^n}\left(\prod_{i=1}^{n} R'_i\right)^{\alpha'-1} e^{-\beta'\sum_{i=1}^{n} R'_i}}{\text{Denom.}} \tag{16}$$

and

$$p(\beta'|D') = \begin{cases} K\beta'^{n\alpha'} e^{-\beta'\sum_{i=1}^{n} R'_i} & a \leq \beta' \leq b \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

where the range a to b is the range defined for the uniform prior probability of β'.

Hence

$$\frac{1}{K} = \int_a^b \beta'^{n\alpha'} e^{-\beta'\sum_{i=1}^{n} R'_i} d\beta' \tag{18}$$

To transform variables, let

$$y = \beta'\sum_{i=1}^{n} R'_i \tag{19}$$

and we find

$$\frac{1}{K} = \frac{1}{\left(\sum_{i=1}^{n} R'_i\right)^{n\alpha'+1}}\left[\gamma\left(n\alpha'+1,\ b\sum_{i=1}^{n} R'_i\right) - \gamma\left(n\alpha'+1,\ a\sum_{i=1}^{n} R'_i\right)\right] \tag{20}$$

where γ is the incomplete gamma function. It is noteworthy that $\sum_{i=1}^{n} R'_i$ is a sufficient statistic for the distribution.

We want the moments of the posterior probability distribution of β'. Firstly,

$$\langle\beta'|D'\rangle = \kappa \int_a^b \beta'^{n\alpha'+1} e^{-\beta'\sum_{i=1}^{n} R'_i} d\beta'$$

$$= \frac{\kappa}{\left(\sum_{i=1}^{n} R'_i\right)^{n\alpha'+2}} \int_{a\Sigma R'_i}^{b\Sigma R'_i} y^{n\alpha'+1} e^{-y}\, dy$$

$$= \frac{\kappa}{\left(\sum_{i=1}^{n} R'_i\right)^{n\alpha'+2}} \left[\gamma(n\alpha'+2,\ b\Sigma R'_i) - \gamma(n\alpha'+2,\ a\Sigma R'_i)\right] \tag{21}$$

multiplying (21) by κ from (20) we get

$$\langle\beta'|D'\rangle = \frac{1}{\left(\sum_{i=1}^{n} R'_i\right)} \frac{\gamma(n\alpha'+2,\ b\Sigma R'_i) - \gamma(n\alpha'+2,\ a\Sigma R'_i)}{\gamma(n\alpha'+1,\ b\Sigma R'_i) - \gamma(n\alpha'+1,\ a\Sigma R'_i)} \tag{22}$$

or for the m^{th} moment of the distribution

$$\langle\beta'^m|D'\rangle = \frac{1}{\left(\sum_{i=1}^{n} R'_i\right)^m} \frac{\gamma(n\alpha'+m+1,\ b\Sigma R'_i) - \gamma(n\alpha'+m+1,\ a\Sigma R'_i)}{\gamma(n\alpha'+1,\ b\Sigma R'_i) - \gamma(n\alpha'+1,\ a\Sigma R'_i)} \tag{23}$$

We have developed computer programs which compute the posterior expected values of β' and its moments. They also compute and plot the posterior probability density distribution for β', including the normalizing factor.[4] This function is a truncated gamma distribution, as will be illustrated.

Applying the analysis to the transformed single cloud data, we obtain the results shown in figure 4. In figure 4a, with the prior

[4]Available on request from EML.

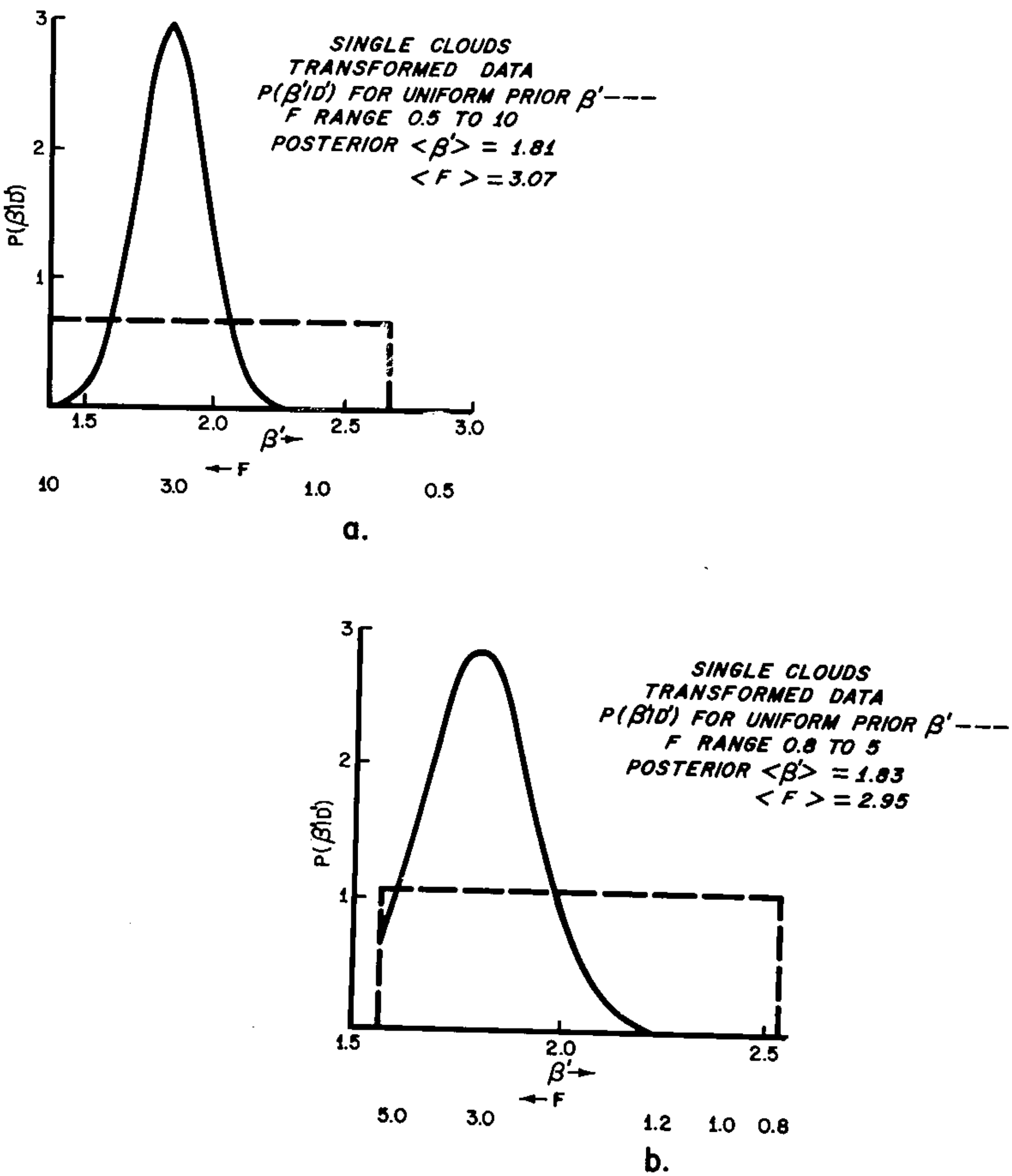

Figure 4. Probability distributions of β' when the prior probability distribution (dashed) is assumed to be uniform. a. Prior probability of β' extends from values corresponding to F = 0.5 to values corresponding to F = 10. b. Prior probability of β' extends from values corresponding to F = 0.8 to values corresponding to F = 5.0.

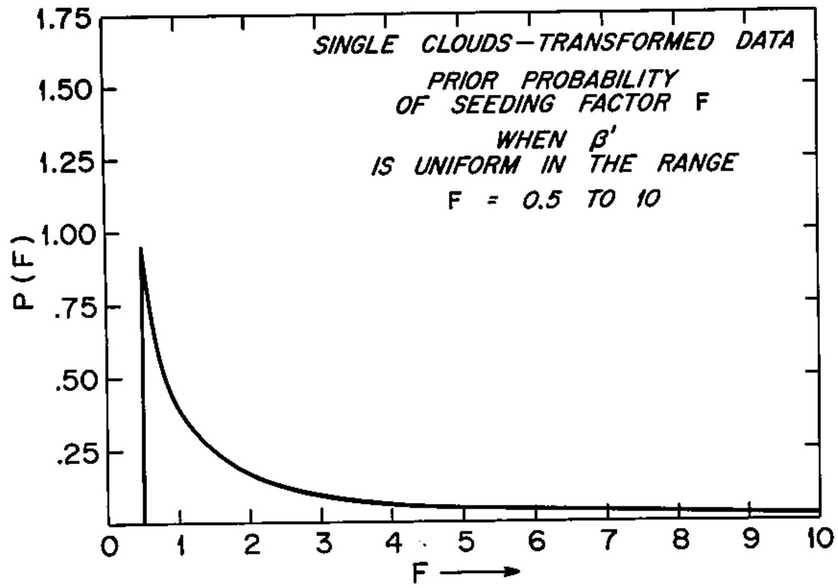

Figure 4c. Prior probability distribution of seeding factor F corresponding to figure 4a.

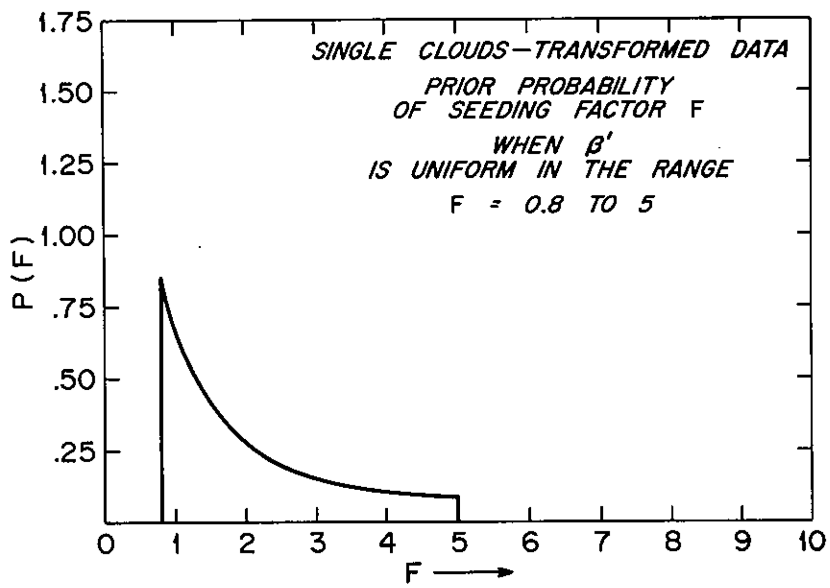

Figure 4d. Prior probability distribution of seeding factor F corresponding to figure 4b.

range of uniform β' corresponding to a seeding factor in the range 0.5 to 10, we see a very sharp peak in the posterior β' probability distribution, corresponding to a posterior expected seeding factor of 3.07. Furthermore, there is virtually no probability that the seeding factor lies below 1.2 or above 8. Reducing the range of the uniform prior β' to an F in the range of 0.8 to 5 (fig. 4b) we find only a small reduction on the posterior expected F, and negligible probability that F is less that 1.2. Figure 4c and 4d show the prior distribution of F corresponding to the two cases of uniform prior on β' that we have treated. Note that neither one is a very favorable prior for F.

The agreement of these results with our earlier ones, together with the somewhat greater information gained from the Bayesian approach, are encouraging. However, the question must be addressed as to whether using the fourth root transformation might have lost any information and/or introduced any bias. The results of the bias test are shown in table 5 and figure 5.

Table 5. Bias Test on Fourth Root Transformation.

EML Single Cloud Control Data

X	$\overline{R}$	$\overline{T}$	$\overline{T}^4$	$\overline{T}^4/\overline{T}^4\,(x=1.0)$
0.2	32.92	1.96177	14.81127	0.21084
0.5	82.29	2.466797	37.02829	0.52710
0.6	98.75	2.58183	44.43350	0.63252
0.8	131.67	2.77436	59.24494	0.84336
1.0	164.59	2.89507	70.24837	1
1.1	181.05	3.00427	81.46214	1.15963
1.2	197.51	3.07034	88.86809	1.26505
1.5	246.88	3.24649	111.08522	1.58132
2.0	329.18	3.48858	148.11353	2.10843
2.5	411.47	3.68872	185.14107	2.63562
3.0	493.76	3.86075	222.17067	3.16264
3.5	576.06	4.01243	259.19694	3.68972
4.0	658.35	4.14864	296.22588	4.21684

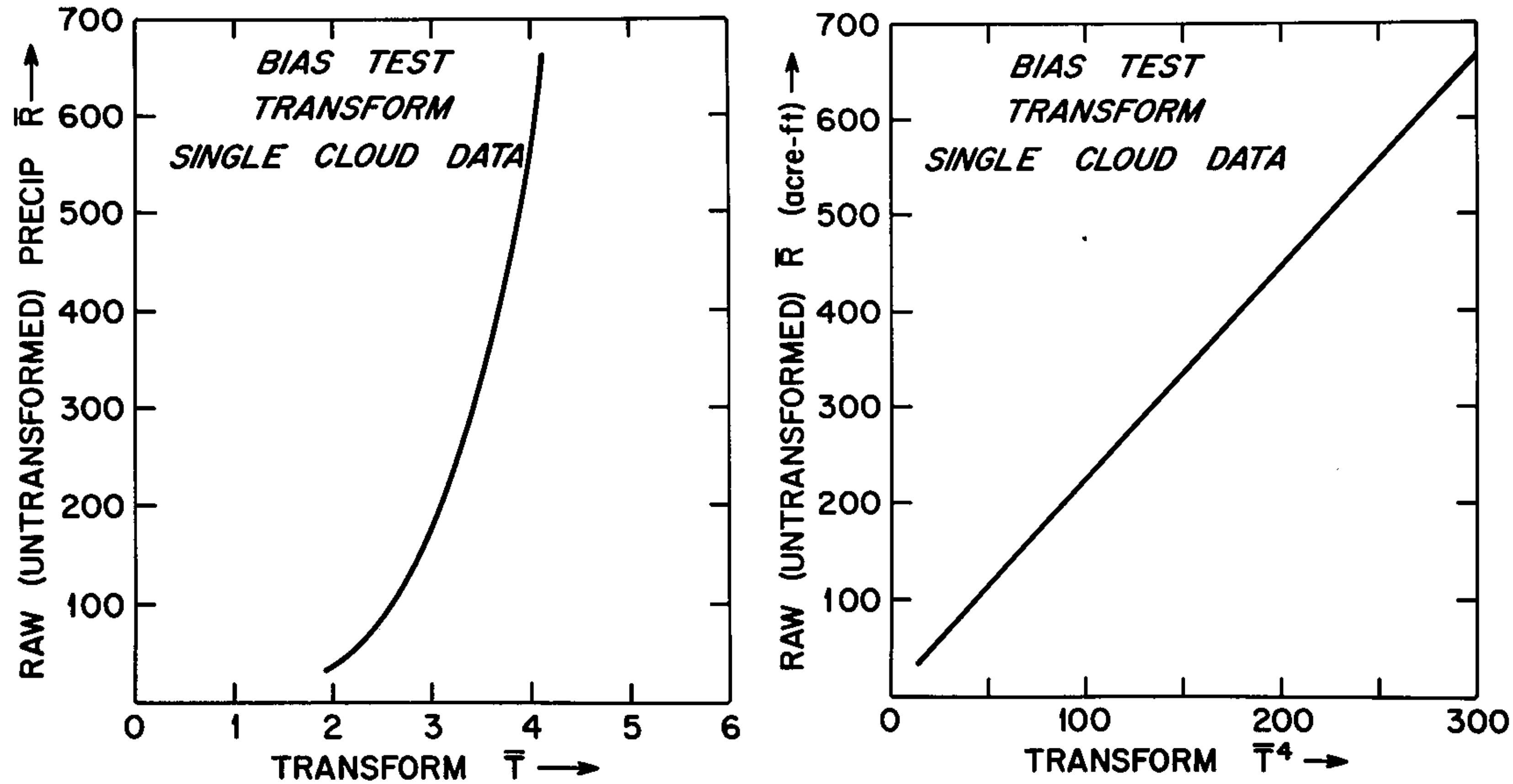

Figure 5. Results of bias test on fourth root data transformation — EML single cloud rainfall data. (left) Untransformed rainfall R in acre-feet plotted against the transform $\overline{T}$. (right) Untransformed rainfall R in acre-feet plotted against $\overline{T}^4$. Note near-linearity.

The bias test consisted of multiplying every cloud rainfall by the factor x and then averaging the 26 cases, presented in the column headed $\overline{R}$. Then the fourth root of each data bit times x was taken and the average of this set presented in the column under $\overline{T}$. Then each entry in the $\overline{T}$ column was raised to the fourth power. The last column presents the ratio of each $\overline{T}^4$ to the value of $\overline{T}^4$ when x = 1. This column tells us, for example, that if seeding increased the rainfall from each cloud by a factor of 3, and if we used transformed data to deduce the seeding factor (without a careful inverse transformation) we would deduce a seeding factor of 3.16 or a little over five percent too high. Figure 5 shows that with these data the transformation is very nearly linear in the important range of the data and only introduces a small positive bias. Nevertheless, it is desirable to work with raw data if and when possible and this is done next.

4. USE OF THE METHOD WITH THE SINGLE CLOUD RAW DATA

Figure 6 compares the histograms for the raw data with those for the transformed. With the high rainfall tails, it was not at first recognized that the gamma distribution might also be applied to the raw data. Extensive meteorological literature has shown that when enough cases are available, the gamma distribution fits a large class of rainfall data (see Thom 1947, 1951, 1957, 1958, 1968; Thom and Vestal, 1968; Mooley and Crutcher, 1968; Mooley, 1972; Barger, Shaw and Dale, 1959 and the bibliographies in these publications). Hence we applied the program DAMAX to the raw single cloud data to obtain the parameters for the best fit gamma distributions and to compare the fits

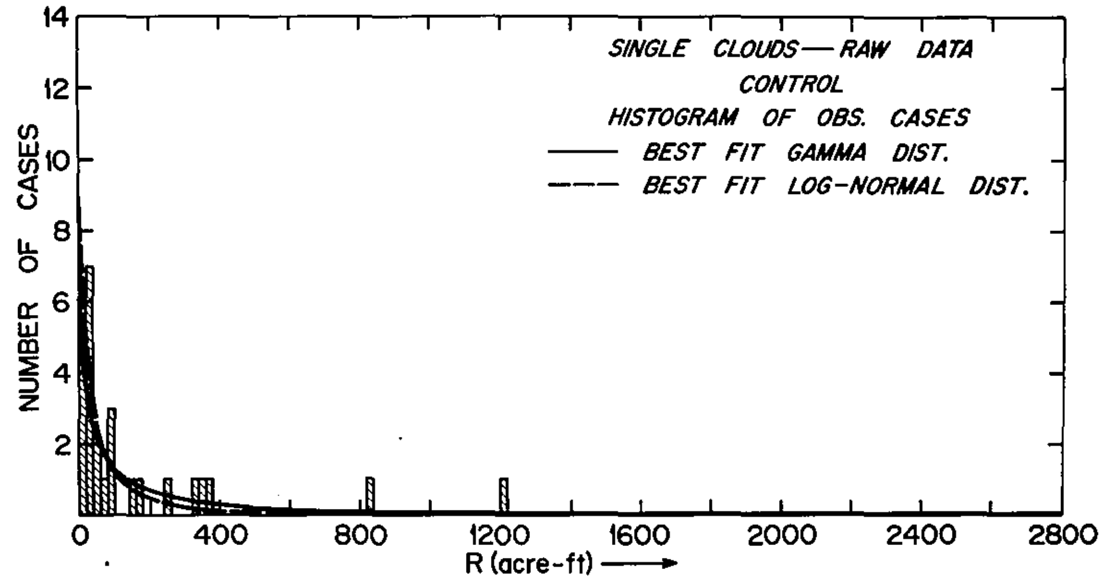

Figure 6. Histograms of single cloud data compared with various classical probability distributions. a. Raw control data versus best-fit gamma and log-normal distributions.

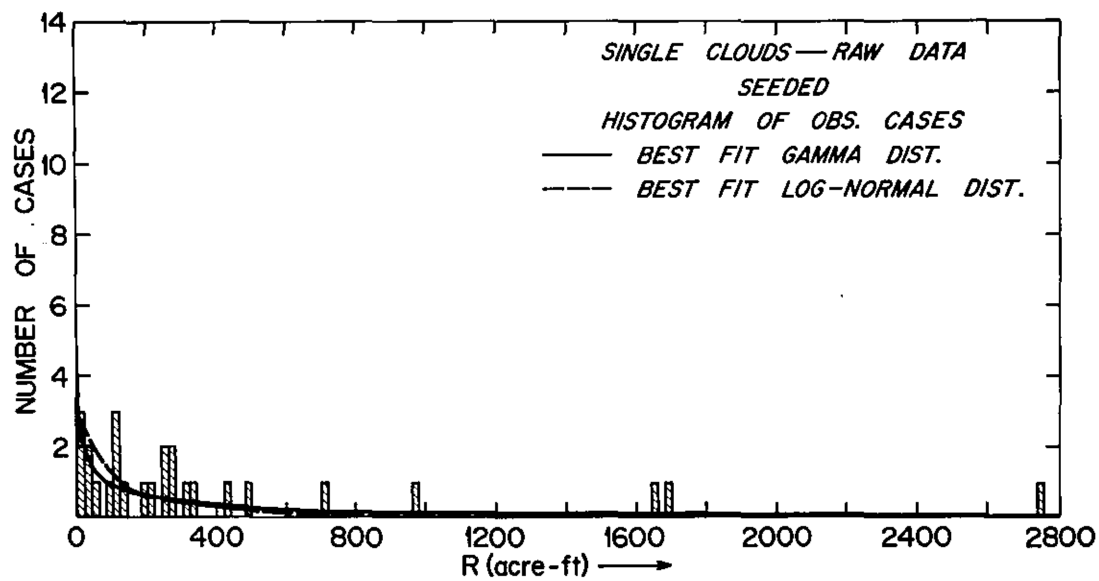

Figure 6b. Raw seeded data versus best-fit gamma and log-normal distributions.

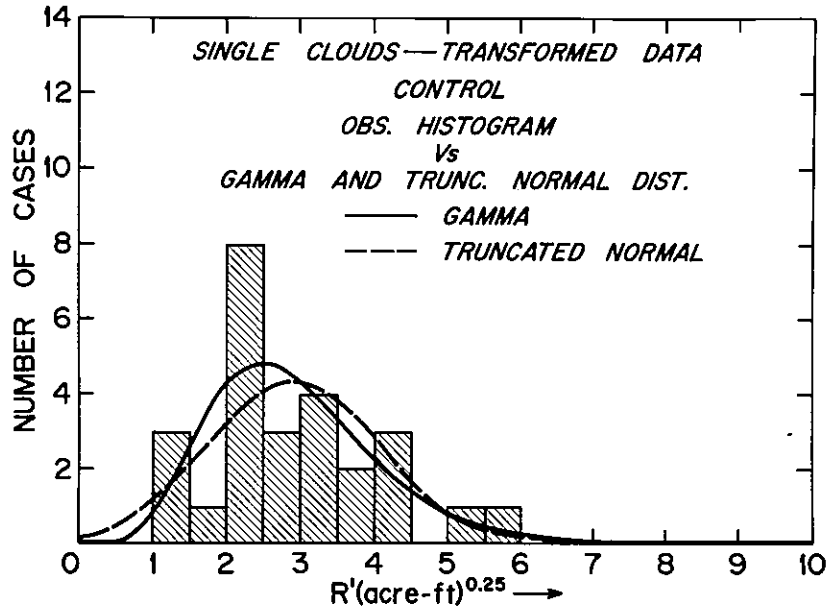

Figure 6c. Transformed (fourth root) control data versus best-fit gamma and truncated normal distribution.

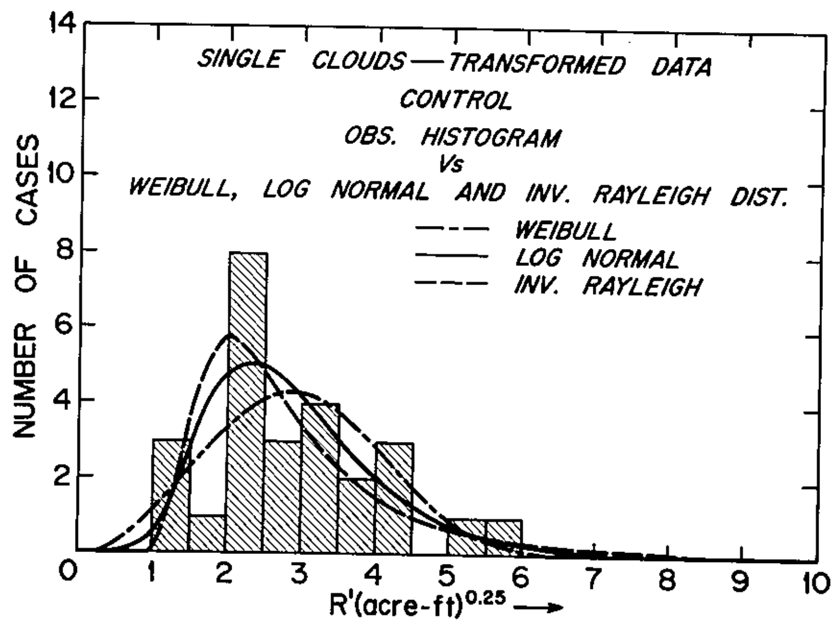

Figure 6d. Transformed (fourth root) control data versus best-fit Weibull, log-normal, and inverse Rayleigh distributions.

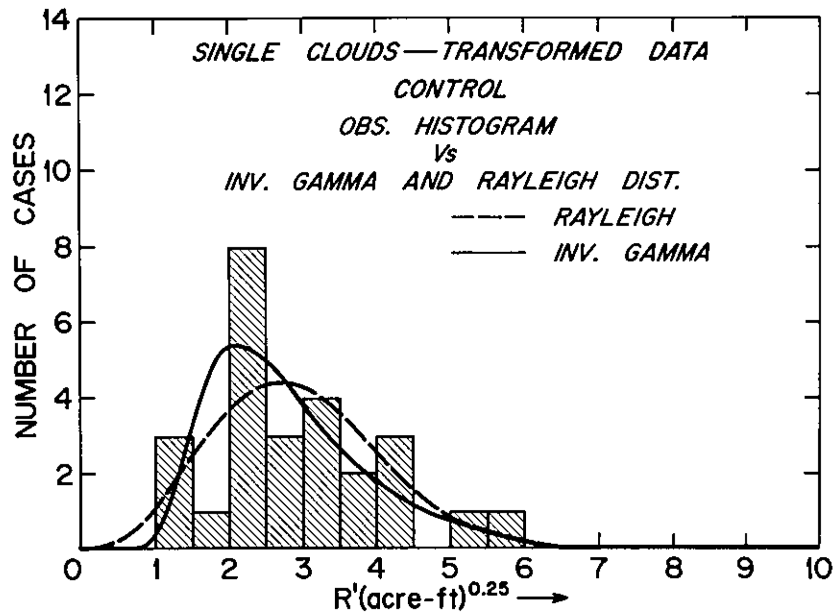

Figure 6e. Transformed (fourth root) control data versus best-fit Rayleigh and inverse gamma distributions.

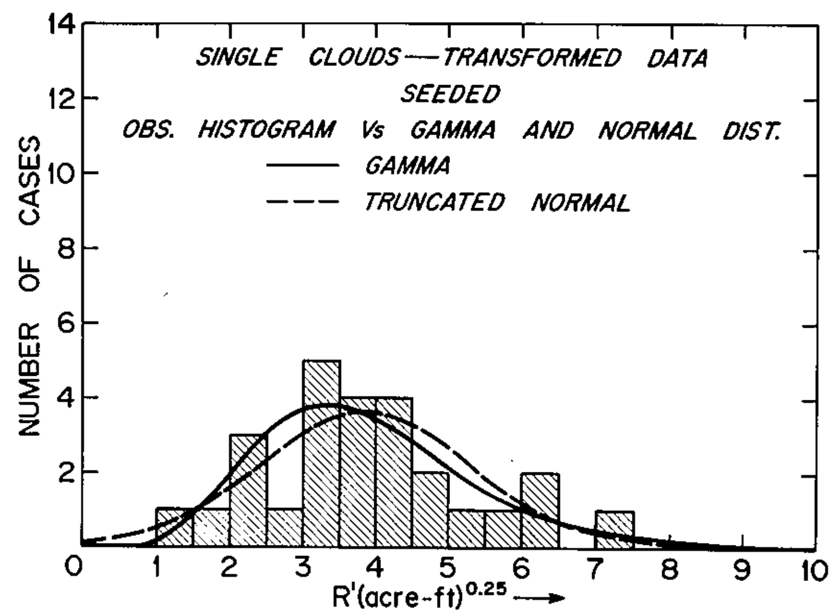

Figure 6f. Transformed (fourth root) seeded data versus best-fit gamma and truncated normal distributions.

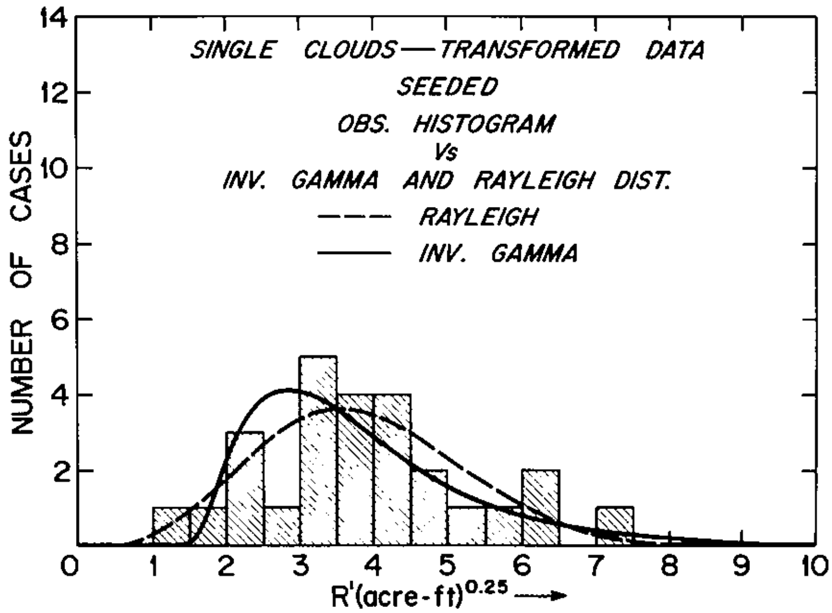

Figure 6g. Transformed (fourth root) seeded data versus best-fit inverse gamma and Rayleigh distributions.

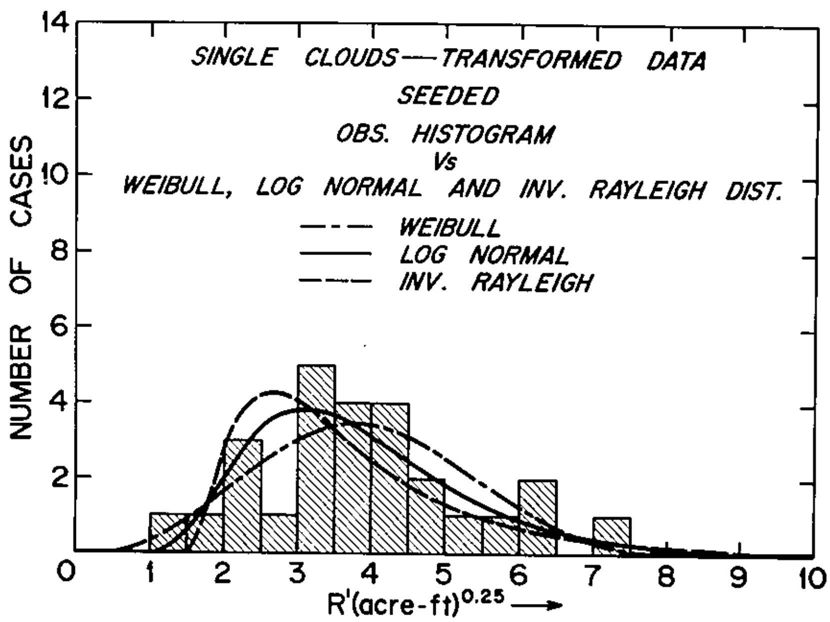

Figure 6h. Transformed (fourth root) seeded data versus best-fit Weibull, log-normal, and inverse Rayleigh distributions.

with the χ^2 calculation[5]. Table 6 presents results for both raw and transformed data.

Table 6.

Fitting of Distribution Functions to EML Single Cloud Data

Distribution	Equation for p(R)
1. Normal Family	$A\ \mathrm{Exp}\ (BR-CR^2)$
2. Log-Normal	$A\ R^B\ \mathrm{Exp}[-C(\log R)^2]$
3. Weibull	$A\ R^B\ \mathrm{Exp}(-CR^{B+1})$
4. Gamma	$A\ R^B\ \mathrm{Exp}(-CR)$
5. Rayleigh	$A\ R^B\ \mathrm{Exp}(-CR^2)$
6. Inverse Gamma	$A\ R^B\ \mathrm{Exp}(-C/R)$
7. Inverse Rayleigh	$A\ R^B\ \mathrm{Exp}(-C/R^2)$

Part I

Control Clouds - Transformed Data

Dist.	-Log(A)	B	C	Rel. Prob.	D.F. = 6 χ^2
1	4.04157	2.02874	0.34878	0.06	6.6
2	3.08051	5.15808	3.08644	0.26	2.3
3	2.34166	1.80100	0.03433	0.09	6.6
4	0.49114	5.52299	2.22360	0.27	4.2
5	2.31008	2.58780	0.18001	0.17	4.9
6	-11.77235	-7.16194	15.37373	0.12	2.5
7	-4.22864	-4.21716	8.40302	0.03	6.0

5 Here the χ^2's have been obtained from the ψ test (see Tribus, 1969, loc. cit., p. 101) to which the χ^2 test is an approximation. For the transformed data slight changes in χ^2 compared to those published earlier (Simpson, 1972) are the result of some minor improvements in the program.

Table 6 (cont'd).

Part II

Seeded Clouds - Transformed Data

Dist.	-Log(A)	B	C	Rel. Prob.	D.F. = 6 χ^2
1	4.86249	1.85362	0.23993	0.14	3.0
2	5.34034	7.34813	3.25198	0.14	2.5
3	3.34546	2.01700	0.01168	0.18	4.7
4	2.47642	6.10433	1.83149	0.24	3.7
5	3.59767	3.00345	0.11728	0.25	2.8
6	-13.84609	-7.29359	20.93732	0.04	5.6
7	-5.17873	-4.23249	14.98414	0.01	7.0

Part III

Control Clouds - Raw Data

Dist.	-Log(A)	B	C	Rel. Prob.	D.F. = 6 χ^2
1	5.10345	-0.00608	0.00000	∿0.	11.7
2	4.46680	0.53952	0.19290	0.66	2.3
3	3.73104	-0.29900	0.03419	0.24	5.9
4	3.64775	-0.43925	0.00341	0.09	3.0
5	3.50278	-0.58723	0.00000	0.01	4.6
6	-0.19194	-1.46183	6.20480	∿0	8.7
7	0.87148	-1.31624	3.71763	∿0	16.3

Part IV

Seeded Clouds - Raw Data

Dist.	-Log(A)	B	C	Rel. Prob.	D.F. = 6 χ^2
1	6.09127	-0.00226	0.00000	0.03	5.9
2	6.72663	1.08703	0.20325	0.30	2.5
3	4.73566	-0.24500	0.01162	0.39	4.7
4	4.52174	-0.36041	0.00145	0.24	6.0
5	4.13570	-0.52839	0.00000	0.04	5.8
6	-0.77388	-1.46878	20.24454	∿0.	8.1
7	0.35567	-1.33478	52.61697	∿0.	28.0

In the case of the gamma distributions, the method of maximum likelihood was also applied to obtain the parameters, with results identical to those presented in table 6. The methods of maximum likelihood and maximum entropy result in the same set of equations to solve when estimating the parameters of the gamma distributions. The concepts of maximum entropy and maximum likelihood are, however, quite different, as explained in Tribus (loc.cit., 1969). The method of Thom's estimators (Shenton and Bowman, 1970) gave identical results with the transformed data and parameters less than two percent higher than those of table 6 with the raw data, where the shape parameter is less than one.

The χ^2's for most cases for most functions are sufficiently low that the null hypothesis cannot be rejected. The gamma distribution appears to fit the transformed data best and perhaps to be a relatively less good fit for the raw seeded data. However, the Monte Carlo experiments in section 6 will show that this conclusion is unwarranted. There we demonstrate that with a sample of this size, it is quite possible that the gamma distribution is the best or nearly perfect fit to the raw data,[6] or anyway no existing evidence militates against our so using it.

Leaving the important questions concerning the determination of distribution functions from small samples to the final section, we proceed next to evaluate the seeding effect from the raw rainfall data.

To apply the same analysis used with the transformed data, with the scale parameter β a measure of the seeding effect, it is necessary to

---

6 We recognize, of course, that if either the raw or the transformed data were perfectly fitted by a gamma distribution, then the other set could be fitted only approximately by that same distribution function.

assume that α, the shape parameter, does not vary between seeded and control populations. With this sample α varies by a little over one percent between seeded and control populations, while β varies by a factor of 2.35. We therefore assume that $\alpha = 0.6$ for both populations. A careful analysis of the consequences of this assumption showed no detectable change in probability for rainfall amounts exceeding ten acre-ft. For rainfall amounts less than ten acre-ft the probability curves for seeded and unseeded rainfalls are brought slightly closer together by the assumption that α is the same for both; hence this assumption will, if anything, cause us to underestimate the seeding effect.

We next construct the table relating F and β, assuming that R for F = 1 is 164.5885 acre-ft, the sample average. Results are presented in table 7.

Table 7. Relation Between Seeding Factor F and β, the Gamma Distribution's Scale Parameter.

F	Single Cloud Raw Data $\alpha = 0.6$	β
0.25		.01456
0.5		.00728
0.8		.00458
1		.00364
1.5		.00243
2		.00182
3		.00121
4		.00091
5		.00072
6		.00060
7		.00052
8		.00040
10		.00036

Table 7 is illustrated graphically in figure 7. There is now a simple inverse relationship between β and F.

As before, we first estimate the seeding factor with the prior β

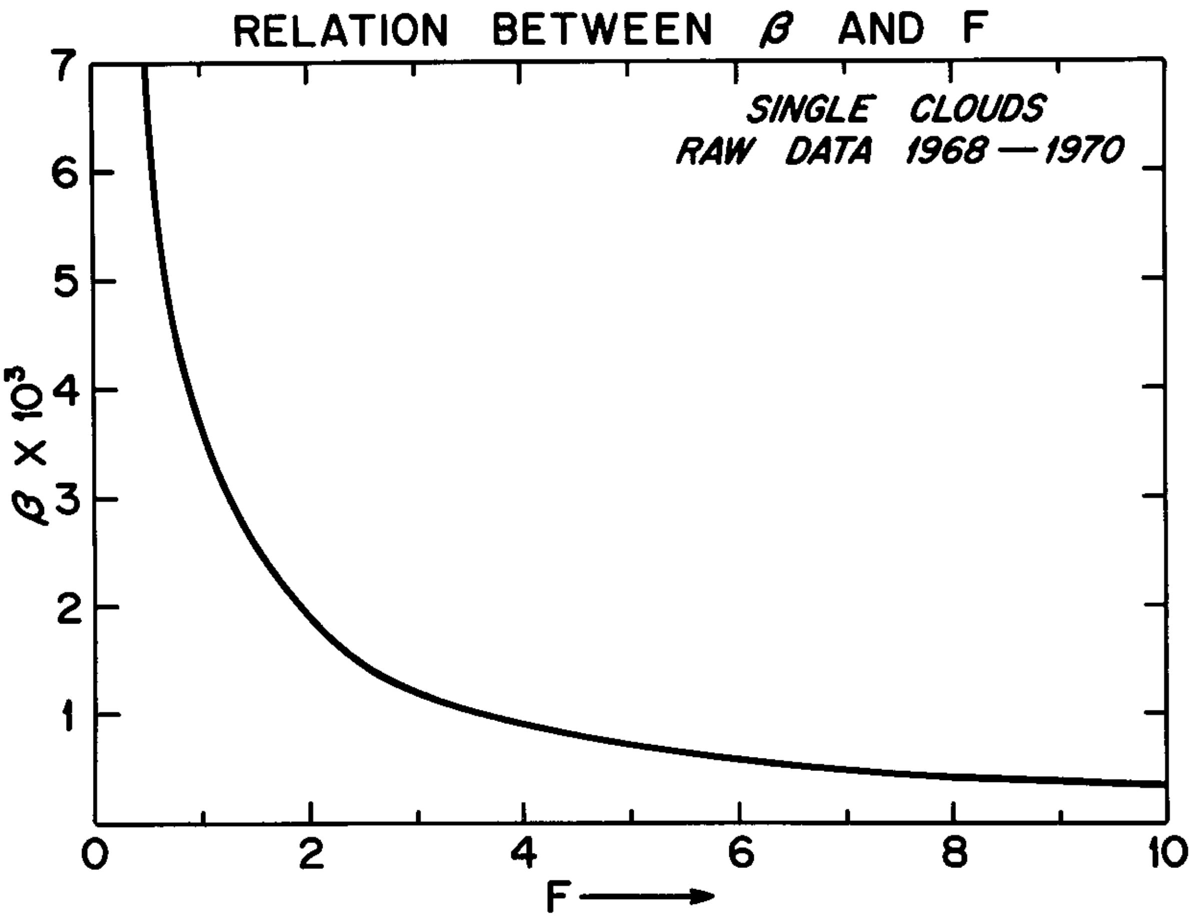

Figure 7. Graph showing relation between β the gamma function scale parameter, and seeding factor F for the single cloud raw data.

a gamma distribution. When n is large, the limiting value of β is .001866, corresponding to an F of 1.96, which is somewhat less than three. Tables 8 and 9 show results corresponding to tables 3 and 4.

Table 8. Raw Single Cloud Data.

Prior β Peaked Gamma Function $K_1 = 12$				Posterior β Gaussian
Prior <β>	F	Post <β>	F	F for 95% Probability
.00728	0.5	.002100	1.80	1.34 - 2.8
.00364	1	.001866	1.96	1.52 - 3.2
.00182	2	.001562	2.35	1.8 - 3.9
.00121	3	.001289	2.85	2.08 - 4.6
.00072	5	.0009802	3.83	2.7 - 5.78

Table 9. Raw Single Cloud Data.

Prior β Flat Gamma Function $K_1 = 2$				Posterior β Gaussian
Prior <β>	F	Post <β>	F	F for 95% Probability
.00728	0.5	.00150	2.4	1.72 - 4.7
.00364	1	.00146	2.5	1.75 - 4.9
.00182	2	.00140	2.6	1.83 - 4.95
.00121	3	.001339	2.7	1.85 - 5.1
.00072	5	.001233	2.9	2.0 - 5.5

As with the transformed data, the flat prior β gives results less sensitive to the prior than does the more peaked curve. Both sets of results, however, give lower seeding factors which are more sensitive to choice of the prior expected value of β. The range for 95 percent is somewhat reduced, however, particularly for the more peaked prior distribution. Figures 8 and 9 illustrate these results. From them, together with figure 7, it is clear that the unfavorable prior distribution is responsible for the lower seeding factor in these cases compared to nearly

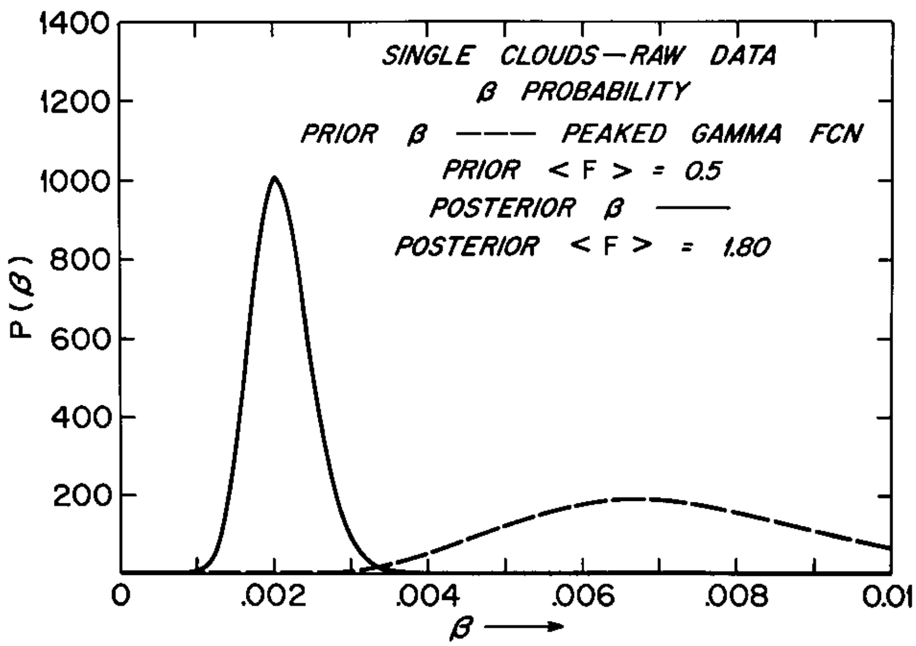

Figure 8. Raw single cloud data - prior (dashed) and posterior (solid) probability distribution of the gamma scale parameter β, when prior β is a gamma distribution with $K_1 = 12$. a. K_2 in prior adjusted so that prior expected β corresponds to $F = 0.5$.

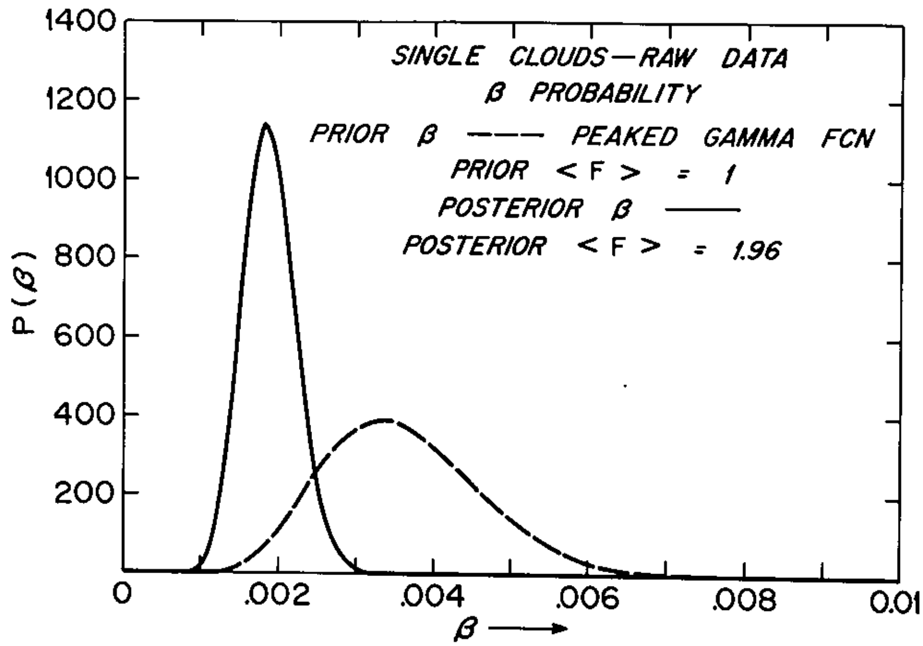

Figure 8b. K_2 *in prior adjusted so that prior expected* β *corresponds to* $F = 1.0$.

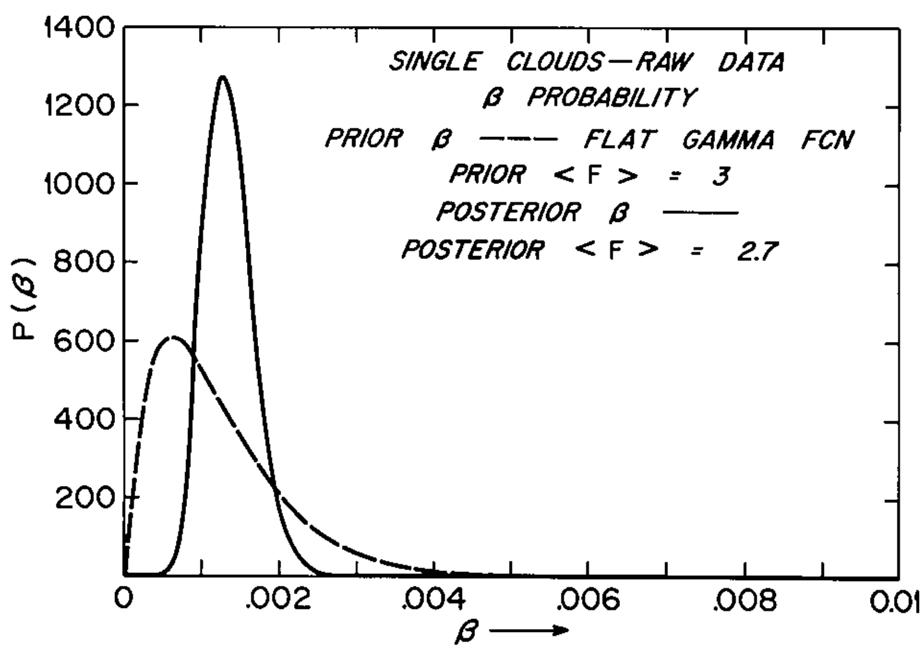

Figure 8c. K_2 *in prior adjusted so that prior expected* β *corresponds to* $F = 3.0$.

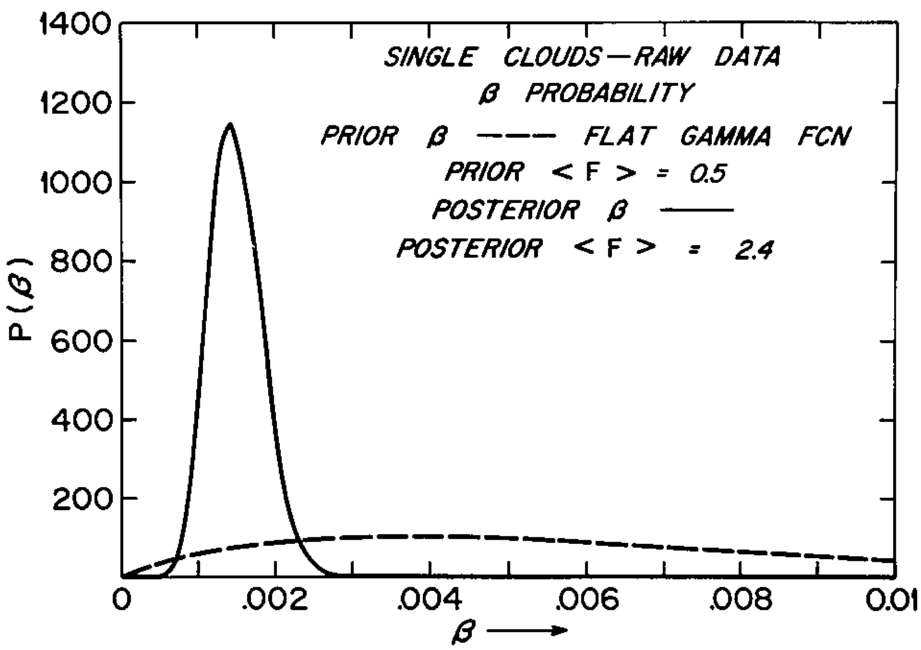

Figure 9. Raw single cloud data - prior (dashed) and posterior (solid) probability distribution of the gamma scale parameter β, when prior β is a gamma distribution with $K_1 = 2$. a. K_2 in prior adjusted so that prior expected β corresponds to $F = 0.5$.

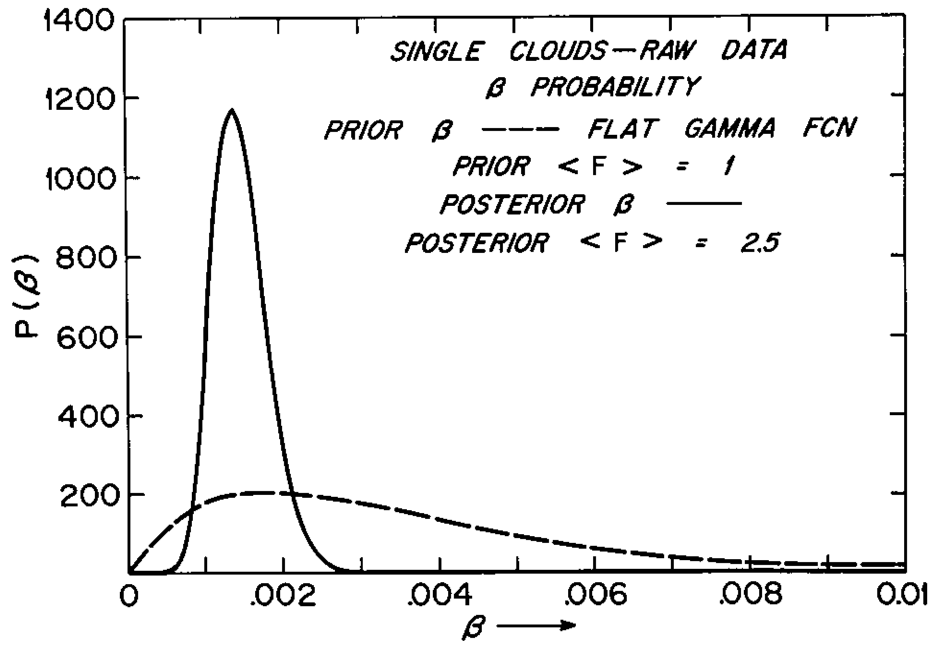

Figure 9b. K_2 *in prior adjusted so that prior expected* β *corresponds to* $F = 1.0$.

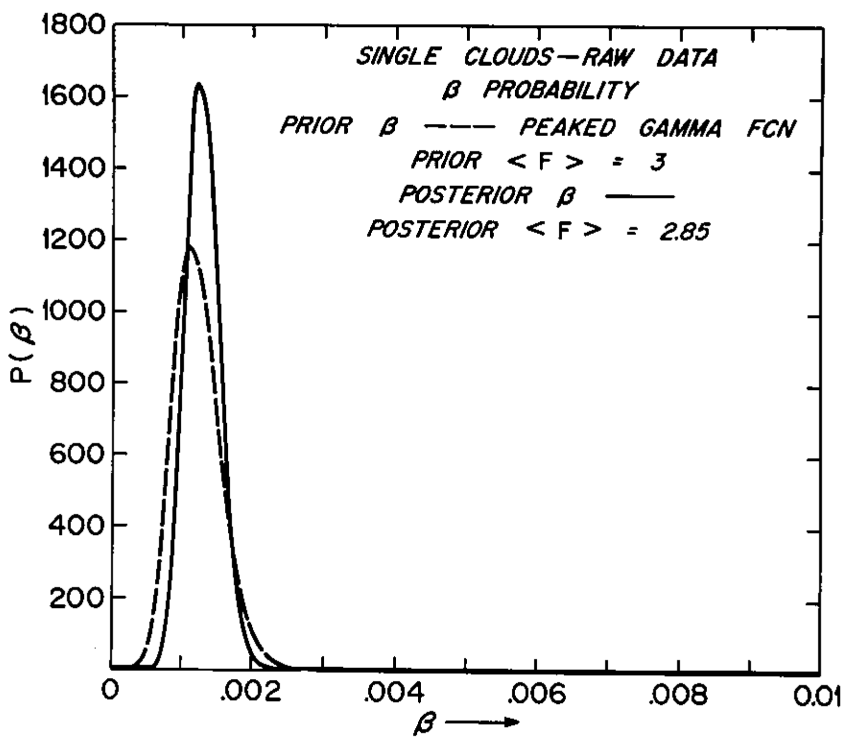

Figure 9c. K_2 *in prior adjusted so that prior expected* β *corresponds to* $F = 3.0$.

all the other methods of its assessment.

We then proceed next to the improved approach, namely a uniform distribution of β over a very wide range of seeding factors. Results are shown in figure 10. With the F range 0.5 to 10 (fig. 10a) we get a very sharp peak in the posterior probability distribution for β, corresponding to an F of 2.5. There is negligible probability that F is less than 1.39. Restricting the prior β range to F from 0.8 to 5 (fig. 10b) does not change the peak value of posterior F. With this prior there is negligible probability that F is less than 1.5. Figures 10c and 10d show the prior probability distributions of F corresponding to the chosen uniform priors of β. These are even more unfavorable than those for the corresponding cases with the transformed data (cf. figs. 4c and 4d). Clearly, the slight reduction on posterior seeding factor that we obtain with the raw data is mainly a result of the less favorable priors, although the data transformation and the assumption of $\alpha = 0.6$ for both seeded and control populations may have contributed slightly. Despites these differences, the data so strongly dominate the posterior distribution of F that it comes out with remarkable consistency in all cases considered so far.

5. ANALYSIS DIRECTLY IN TERMS OF SEEDING FACTOR

5.1 Seeding Factor with Raw Rainfall Data

The optimal approach to seeding factor evaluation is to treat the probability distributions of the seeding factor itself. This method, of course, involves setting the prior probability distribution on seeding factor, for which most people would have a greater intuitive preference than for using its reciprocal or a function of its reciprocal; most important, this procedure permits a diffuse prior probability on the

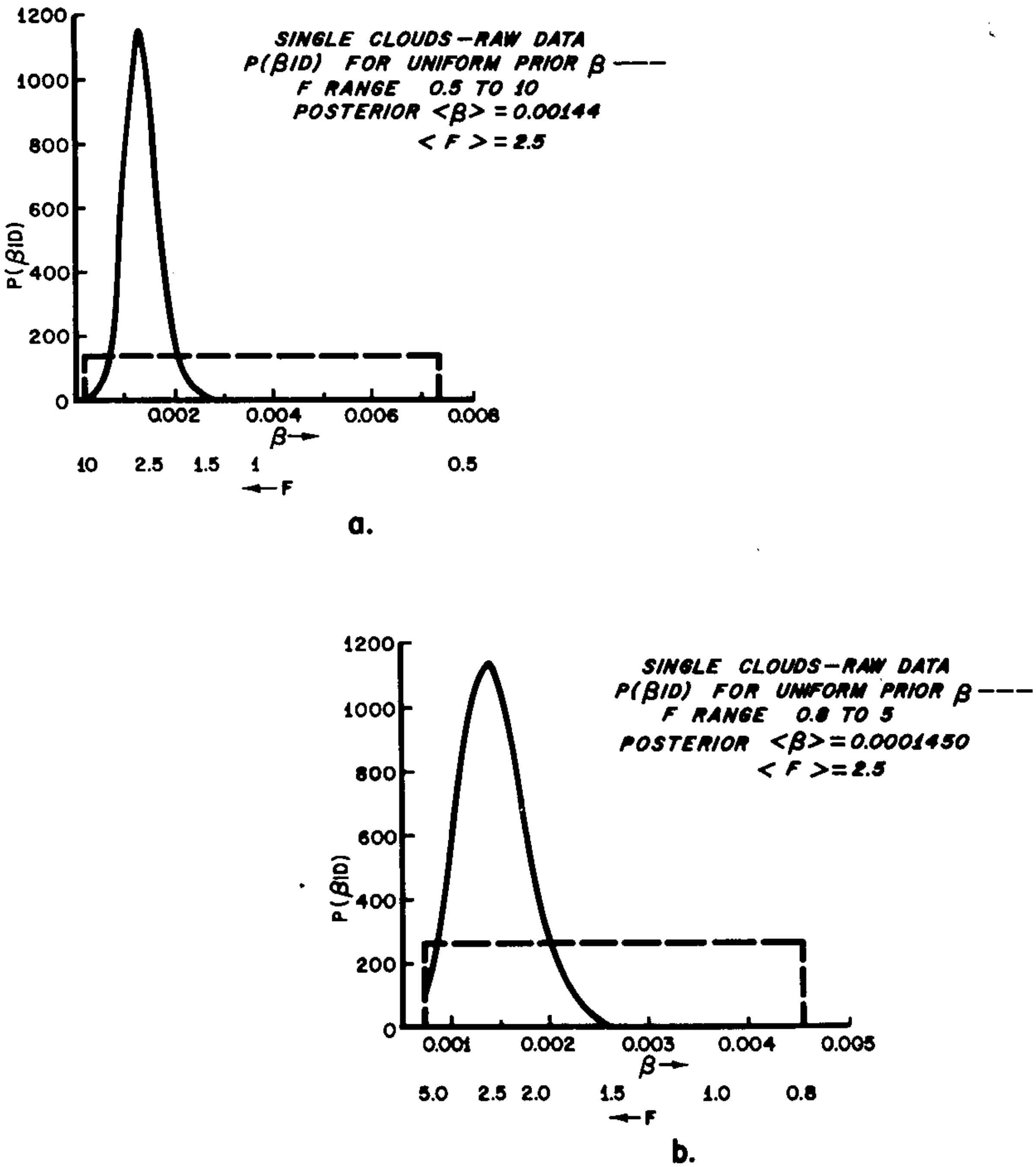

Figure 10. Raw single cloud data - uniform prior (dashed) probability of β. Posterior probability solid. a. Uniform prior in range corresponding to F from 0.5 to 10. b. Uniform prior in range corresponding to F from 0.8 to 5.

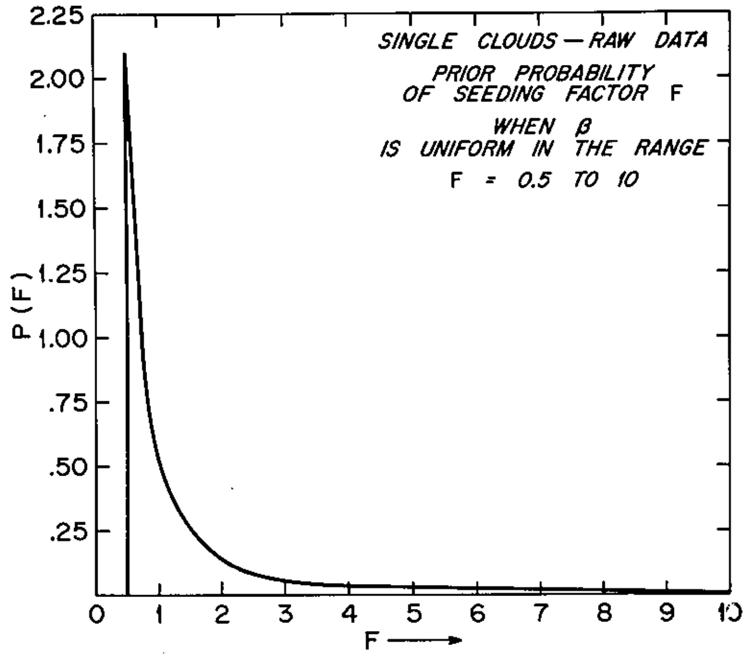

Figure 10c. Prior probability distribution of seeding factor F corresponding to figure 10a.

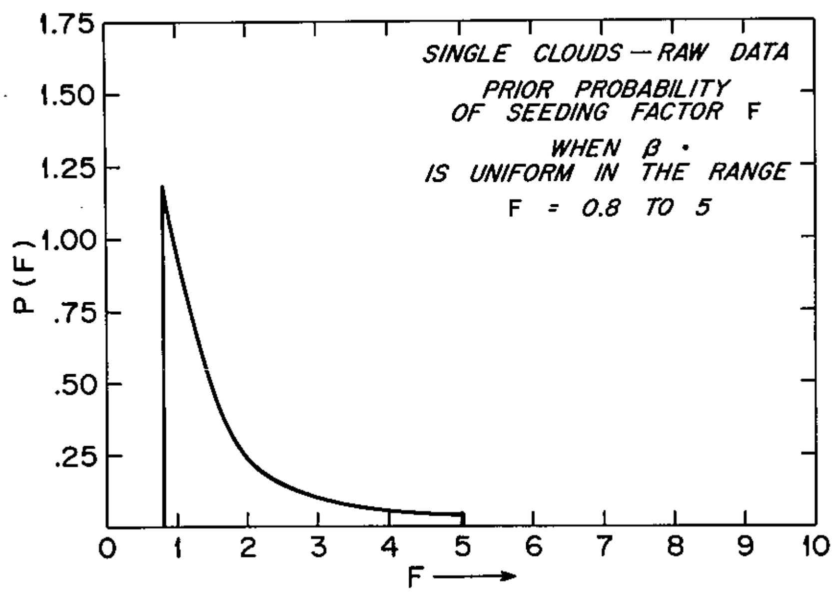

Figure 10d. Prior probability distribution of seeding factor F corresponding to figure 10b.

seeding factor itself, namely uniform over a wide range.

Therefore, in this treatment we will eliminate β altogether from the rainfall probability density distribution and write instead

$$p(R) = \frac{\left(\frac{\alpha}{\langle R\rangle_{NS}\ F}\right)^{\alpha}}{\Gamma(\alpha)} R^{\alpha-1} e^{-\left(\frac{\alpha}{\langle R\rangle_{NS}\ F}\right)R} \qquad (24)$$

where R is now rainfall in acre-ft; $\langle R\rangle_{NS}$ is the sample average or expected value of the unseeded distribution. F is the seeding factor, defined just as before, namely

$$\langle R\rangle = \langle R\rangle_{NS}\ F \qquad (25)$$

and

$$V^2(R) = \frac{1}{\alpha} \qquad (26)$$

Again, we will begin with a prior probability assignment to F which permits an analytic solution, namely an inverse gamma function as follows:

$$p(F) = \frac{(K_2)^{K_1+1}}{\Gamma(K_1+1)} F^{-K_1-2} e^{-K_2/F} \qquad (27)$$

The first two moments of the inverse gamma distribution are

$$\langle F\rangle = \frac{K_2}{K_1}$$

and (28)

$$V^2(F) = \frac{1}{K_1-1}$$

To find the posterior probability distribution for seeding effect F, we apply Bayes equation and proceed as follows:

$$p(F|D) = C\ p(F)\ p(D|F)$$

$$= C F^{-K_1-2} e^{-K_2/F} F^{-n\alpha} \exp\left[-\frac{\alpha \sum_{i=1}^{n} R_i}{\langle R\rangle_{NS} F}\right] \tag{29}$$

$$= C F^{-n\alpha-K_1-2} \exp\left[-\frac{1}{F}\left(K_2 + \frac{\alpha \sum_{i=1}^{n} R_i}{\langle R\rangle_{NS}}\right)\right] \tag{30}$$

where with the single clouds $\langle R\rangle_{NS}$ = 164.588 acre-ft, the average of the 1968-1970 sample. This time we find the normalizing constant C by setting up the integration and transforming to y = 1/F. The resulting integral is recognized as that of a gamma, so that the constant is known to be

$$C = \frac{\left(K_2 + \frac{\alpha\Sigma R_i}{\langle R\rangle_{NS}}\right)^{n\alpha + K_1 + 1}}{\Gamma(n\alpha + K_1 + 1)}$$

Now

$$\langle F|D\rangle = \frac{K_2 + \frac{\alpha \sum_{i=1}^{n} R_i}{\langle R\rangle_{NS}}}{K_1 + n\alpha} \tag{31}$$

and

$$V^2(F|D) = \frac{1}{n\alpha + K_1 - 1} \tag{32}$$

It is noteworthy and logical that when n = 0 (i.e. there are no data) that in (31) $\langle F|D\rangle$ reduces to K_2/K_1. When n becomes very large, the expected value of the seeding factor approaches $\overline{R}_S/\overline{R}_{NS}$ as its coefficient of variation shrinks.

We now consider seeding factor probabilities using the single cloud raw data, (30) - (32) and combinations of K_1 and K_2 as listed in table 10; note that K_1 must exceed zero for finite prior expectation and that it must exceed one for finite prior variance.

Table 10. Seeding Factor - Raw Data - Prior Inverse Gamma Function. Values of K_1, K_2; Prior Expectations and Standard Deviations.

Case No.	K_1	K_2	Prior F	Prior σ
1	2.25	6.75	3	2.68
2	1	3	3	∞
3	1	1	1	∞
4	1	0.5	0.5	∞
5	10	20	2	0.667
6	10	5	0.5	0.167
7	0.5	0.5	1	not defined

Table 11 gives the seeding factor expectation, standard deviation and $\pm$ 2σ range after seeing the seeded single cloud data. As illustrated by the near symmetry of the solid curves of probability density for seeding factor, the $\pm$ 2σ range is a fair approximation to the integrated 95 percent probability range.

Table 11. Seeding Factor - Raw Data - Posterior Probabilities With Prior Inverse Gamma Function.

Case No.	<F>	σ	$\pm$ 2σ range
1	2.72	0.64	1.37 - 3.93
2	2.70	0.66	1.34 - 4.01
3	2.58	0.63	1.31 - 3.85
4	2.55	0.63	1.30 - 3.80
5	2.42	0.48	1.46 - 3.37
6	1.83	0.36	1.11 - 2.55
7	2.63	0.66	1.32 - 3.94

The seven cases are illustrated graphically in figures 11 and 12. Cases 4 and 6 show what happens even with a prior prejudice against the seeding which is so extreme that virtually no probability favoring even a positive effect is permitted (fig. 12a and particularly fig. 12c). Even with this prejudice, the data overcome it and all but two of the posterior probabilities exceed 2.5. With an extremely small exception in case 5, negligible posterior probability of a negative seeding effect remains after taking the data the data into account. However, it is still better to take a more diffuse prior probability distribution for the seeding factor.

The final effort for the raw data is to take a uniform prior probability, namely

$$p(F) = \frac{1}{b-a} \text{ for } a \leq F \leq b \tag{33}$$

so that the posterior probability density distribution for seeding factor F is

$$\begin{aligned} p(F|D) &= C_1 \; p(D|F) \quad \text{for } a \leq F \leq b \\ &= 0 \qquad \text{elsewhere} \\ &= C_1 \; F^{-n\alpha} \exp\left[- \frac{\alpha \sum_{i=1}^{n} R_i}{\langle R \rangle_{NS} \; F} \right] \text{ for } a \leq F \leq b \\ &= 0 \qquad \text{elsewhere} \end{aligned} \tag{34}$$

The normalizing constant C_1 may either be found by adapting the computer program described, setting the integral from a to b equal to one, or analytically as follows:

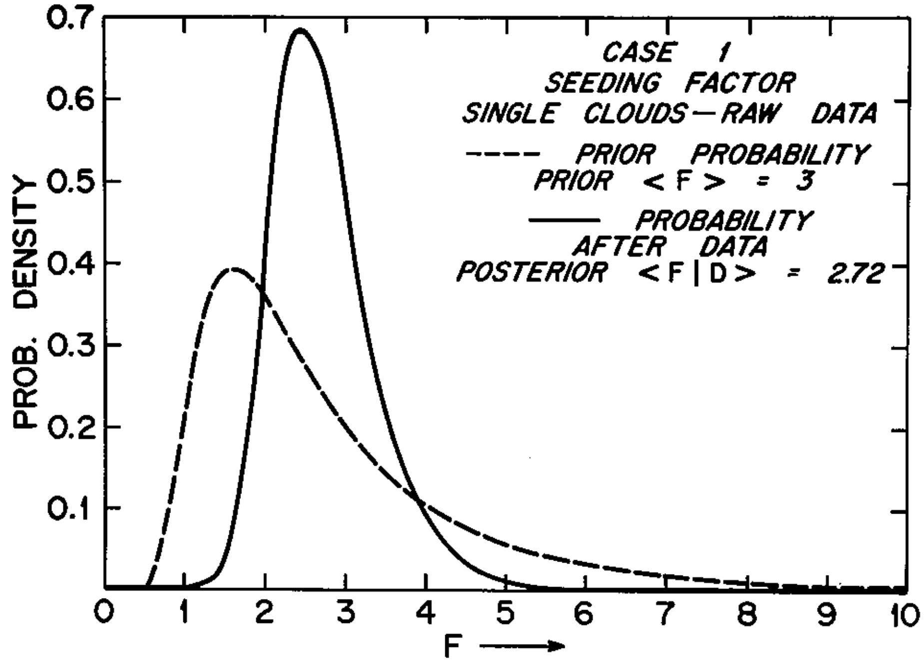

Figure 11. Direct seeding factor analysis - raw single cloud data. Prior probability on seeding factor (dashed) is an inverse gamma function. Posterior probability on seeding factor solid. a. Case 1. Prior expected F = 3.

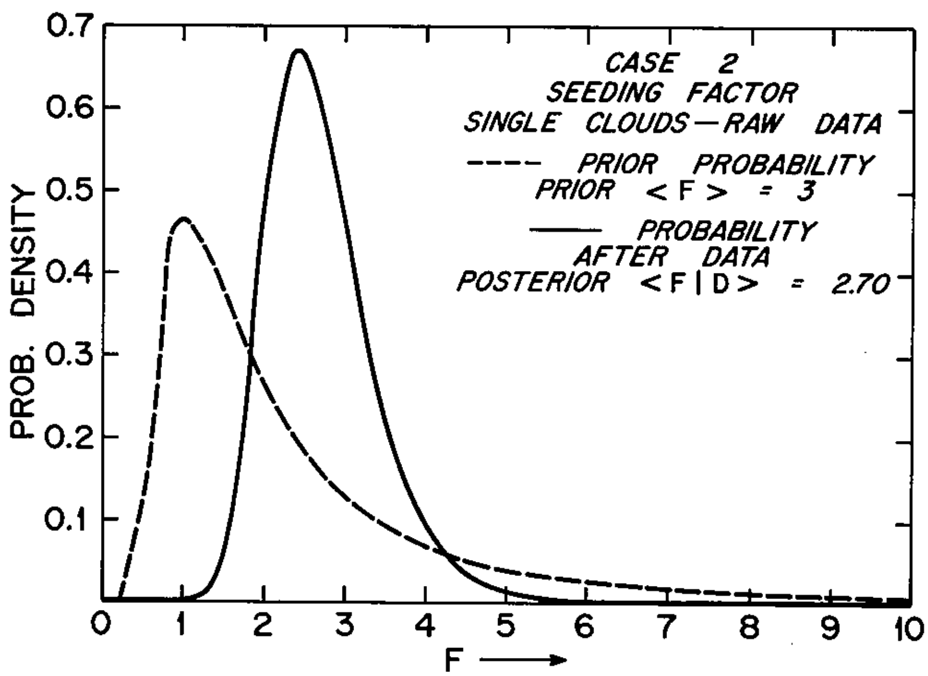

Figure 11b. Case 2. Prior expected F = 3.

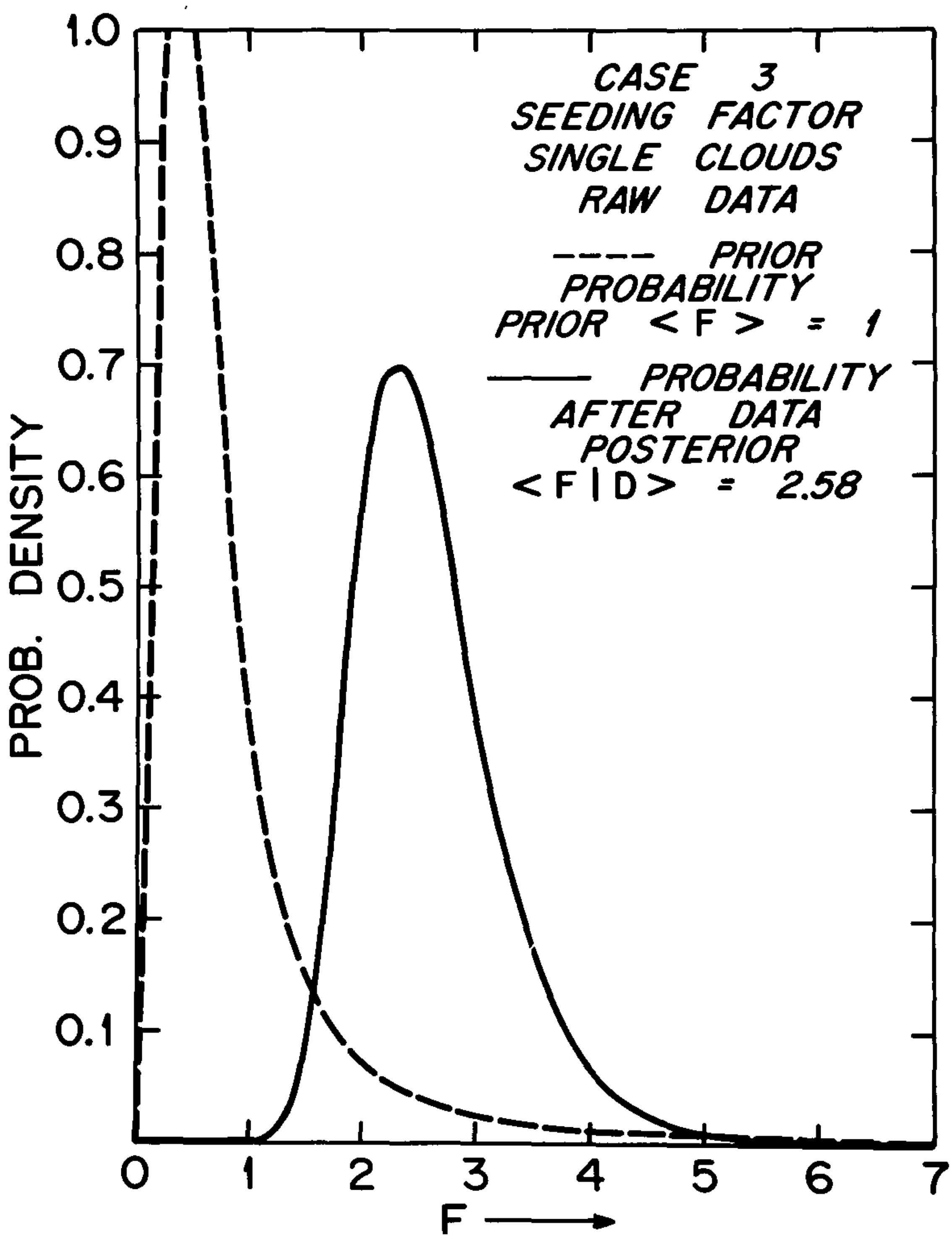

Figure 11c. Case 3. Prior expected F = 1.

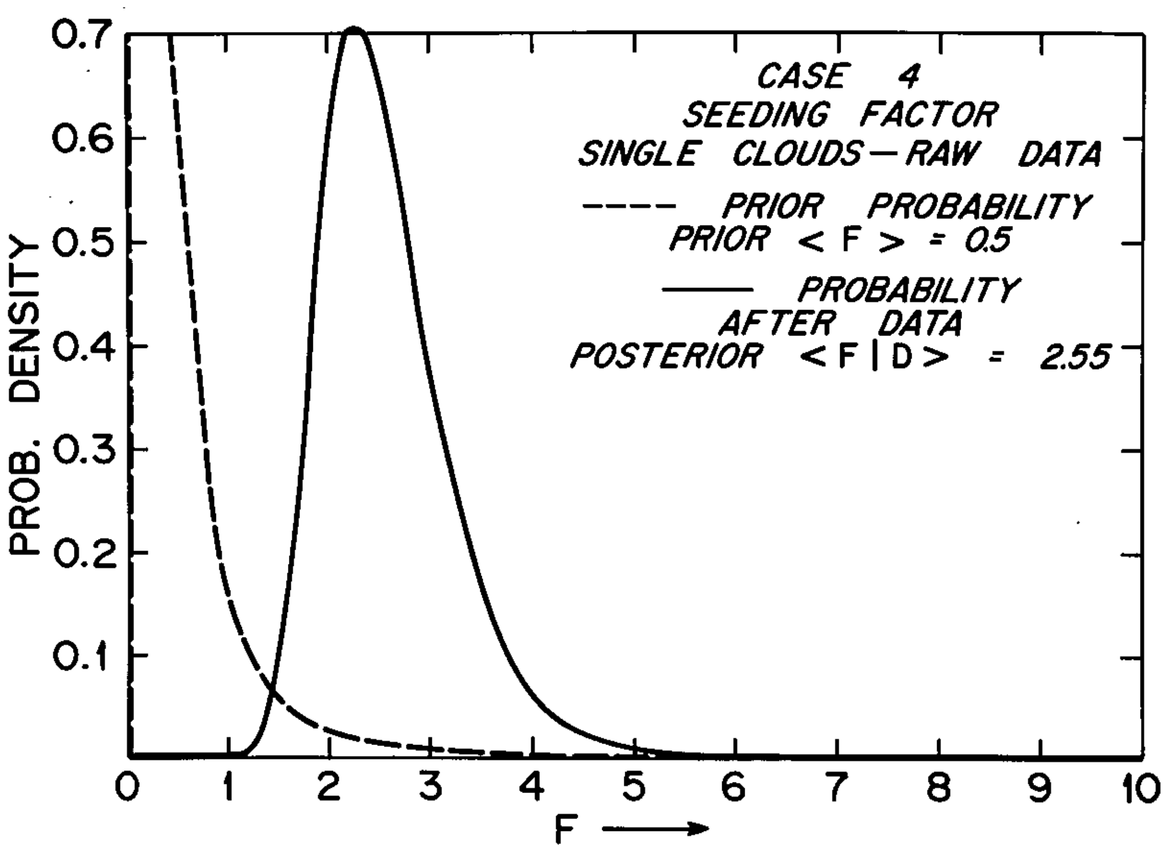

Figure 12. Same as figure 11 with different priors. a. Case 4. Prior expected F = 0.5.

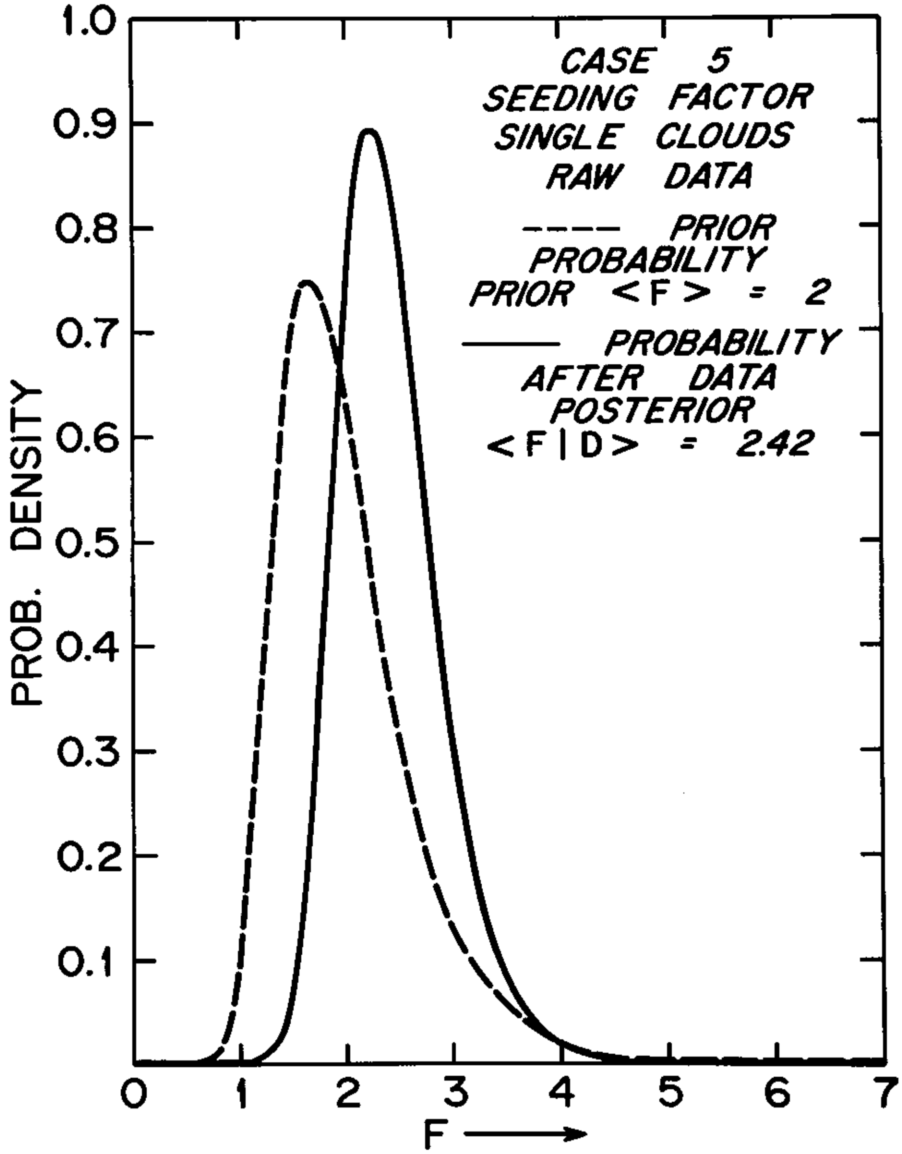

Figure 12b. Case 5. Prior expected F = 2.

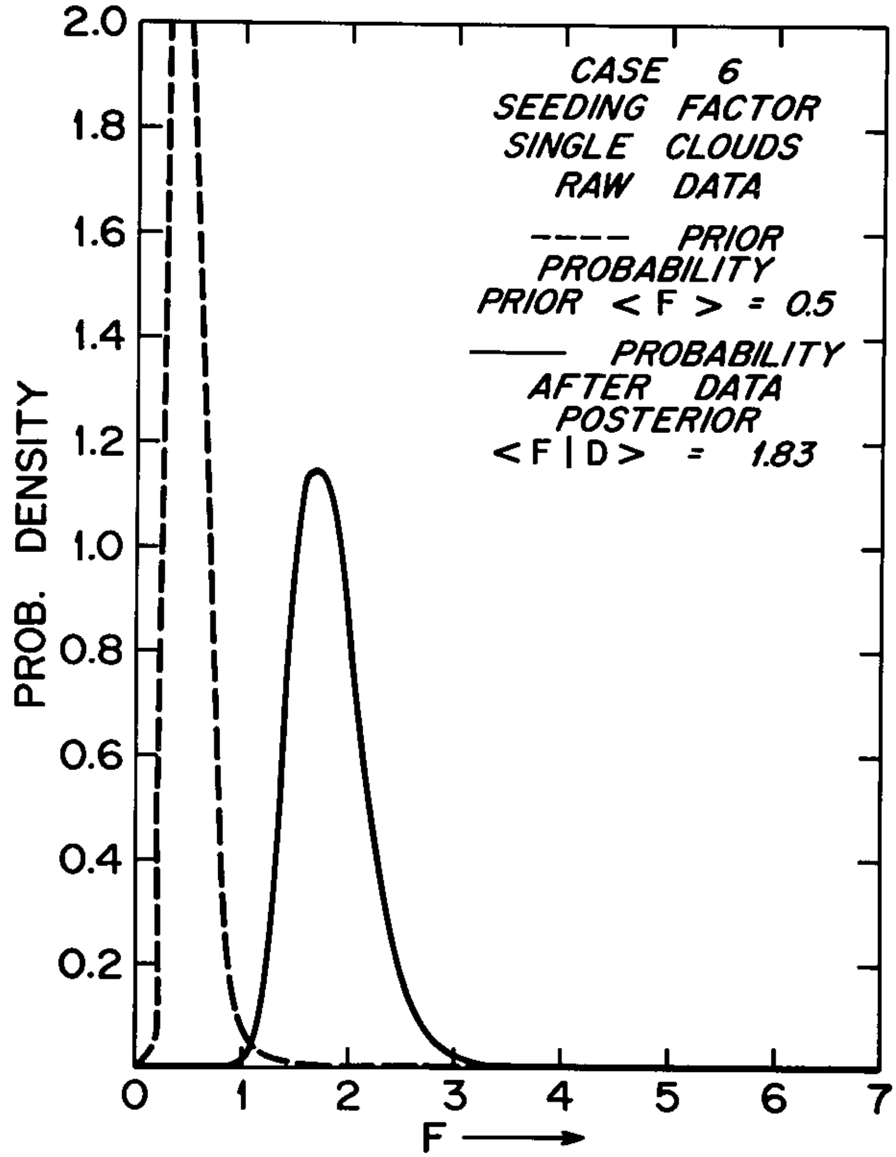

Figure 12c. Case 6. Prior expected F = 0.5.

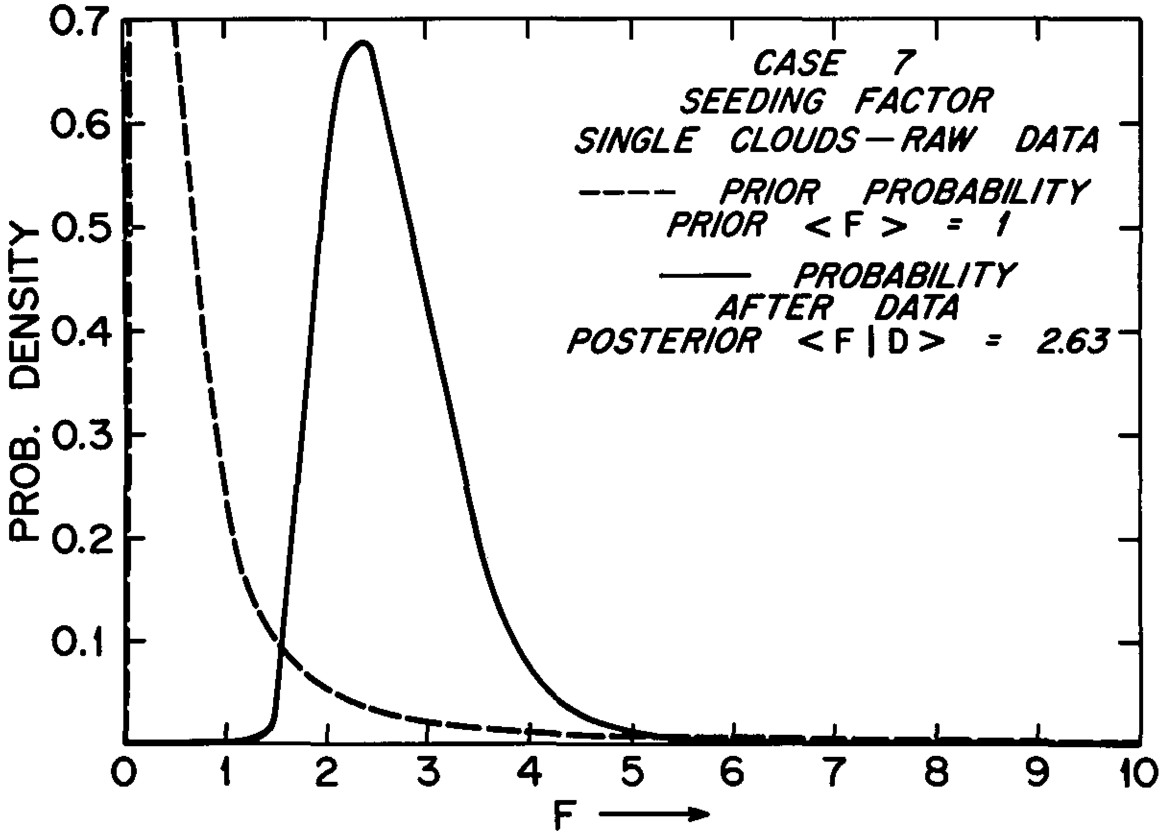

Figure 12d. Case 7. Prior expected F = 1.

$$C_1 = \left(\frac{\alpha \sum_{i=1}^{n} R_i}{\langle R_{NS}\rangle}\right)^{n\alpha-1} \Bigg/ \left[\gamma\left(n\alpha-1, \frac{\alpha \sum_{i=1}^{n} R_i}{\langle R_{NS}\rangle a}\right) - \gamma\left(n\alpha-1, \frac{\alpha \sum_{i=1}^{n} R_i}{\langle R_{NS}\rangle b}\right)\right] \quad (35)$$

To simplify, let $\mu = \dfrac{\alpha \sum_{i=1}^{n} R_i}{\langle R_{NS}\rangle}$ then

$$C_1 = \mu^{n\alpha-1} \Big/ [\gamma(n\alpha-1,\mu/a) - \gamma(n\alpha-1,\mu/b)]$$

$$\langle F\rangle = \mu \frac{\gamma(n\alpha-2,\mu/a) - \gamma(n\alpha-2,\mu/b)}{\gamma(n\alpha-1,\mu/a) - \gamma(n\alpha-1,\mu/b)} \quad (36)$$

$$\langle F^2\rangle = \mu^2 \frac{\gamma(n\alpha-3,\mu/a) - \gamma(n\alpha-3,\mu/b)}{\gamma(n\alpha-1,\mu/a) - \gamma(n\alpha-1,\mu/b)}$$

Here γ is the incomplete gamma function as before.[7]

Two ranges of uniform prior probability on seeding factor were considered, namely a reasonable range from F = 0.8 to 5 and an extreme range from F = 0.5 to 10. Results are shown in table 12 and figure 13a and b.

Table 12. Single Clouds - Raw Data. Uniform Prior Probability on Seeding Factor.

Case 1 - Reasonable Range 0.8 - 5	Case 2 - Extreme Range 0.5 - 10
After data:	After data:
$\langle F\|D\rangle = 2.99$	$\langle F\|D\rangle = 3.08$
$\sigma(F\|D) = 0.72$	$\sigma(F\|D) = 0.87$

[7] When using the normalized incomplete gamma function it is necessary to multiply $\langle F\rangle$ by $1/(n\alpha-2)$ and $\langle F^2\rangle$ by $1/[(n\alpha-2)(n\alpha-3)]$.

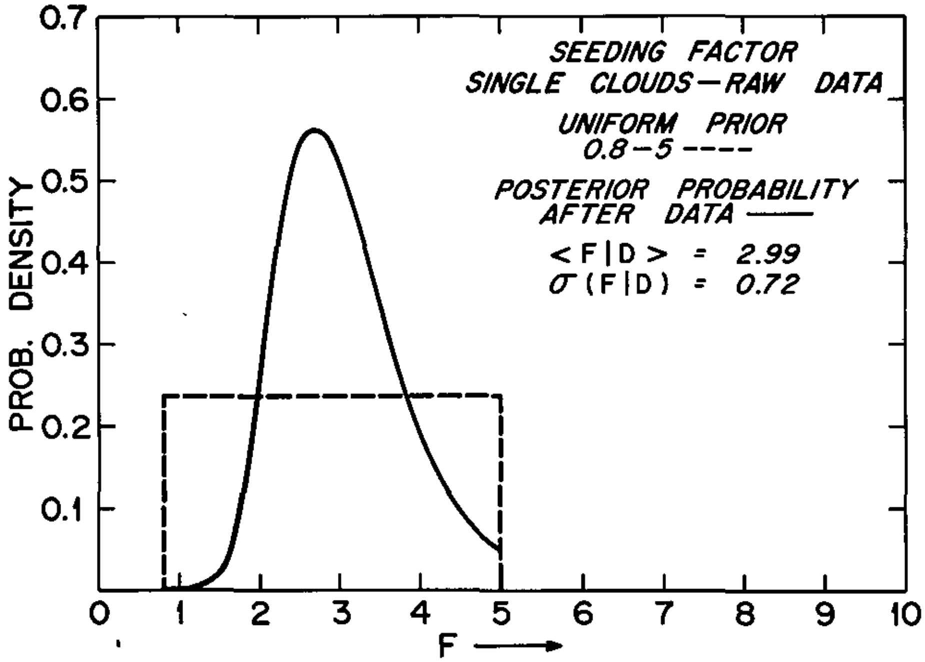

Figure 13. Direct seeding factor analysis - raw single cloud data. Uniform prior probability on seeding factor (dashed). Posterior probability on seeding factor (solid). a. Reasonable range on prior; F from 0.8 to 5.

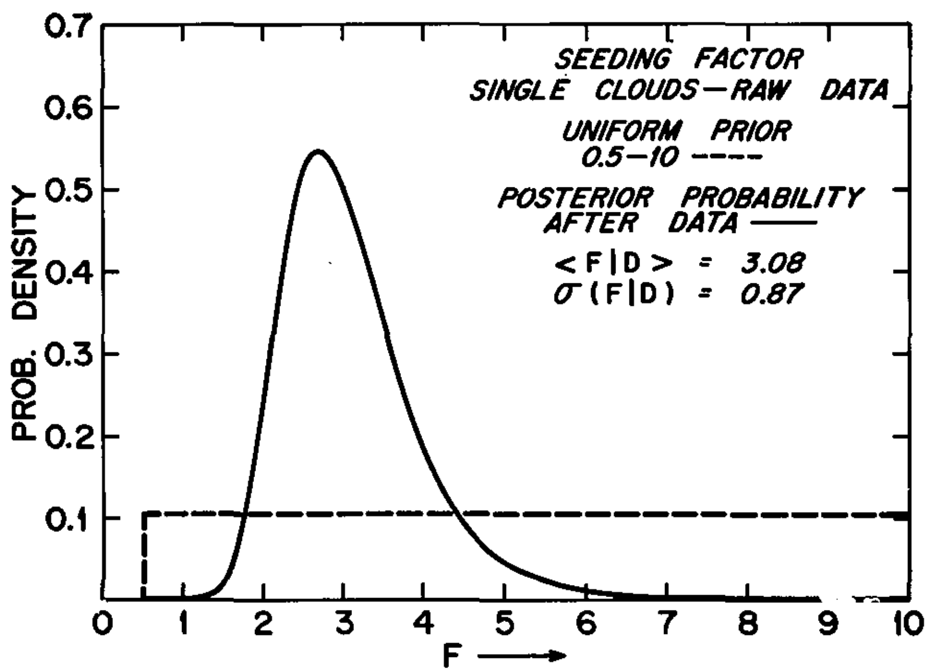

Figure 13b. Extreme range on prior; F from 0.5 to 10.

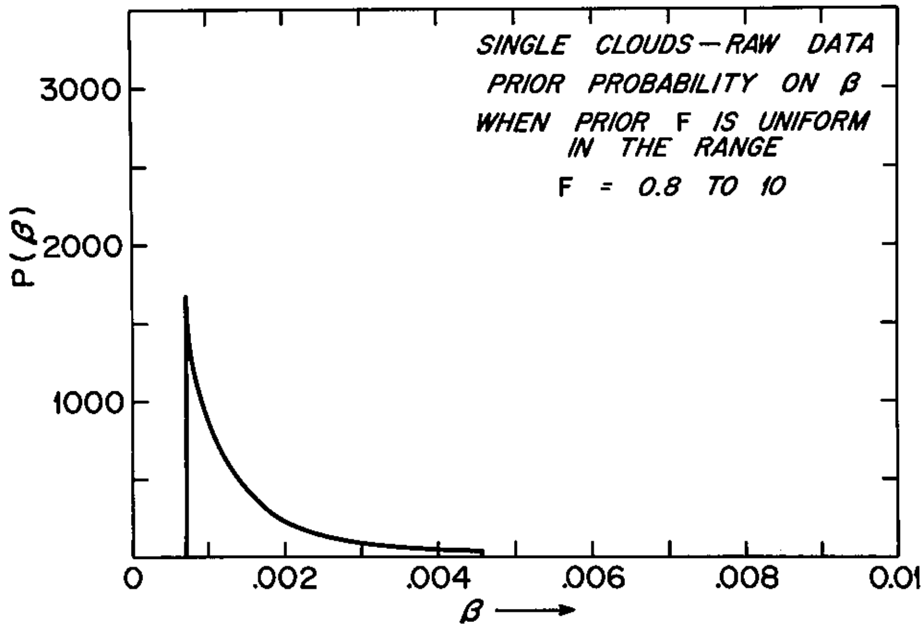

Figure 13c. Prior probability distribution of gamma function scale parameter β, corresponding to figure 13a.

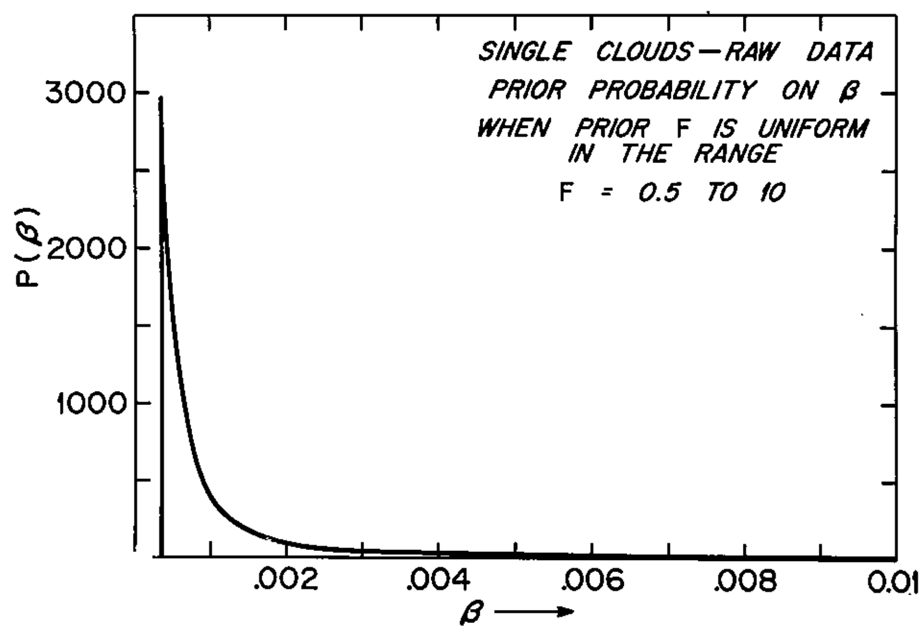

Figure 13d. Prior probability distribution of gamma function scale parameter β, corresponding to figure 13b.

Again, the most likely seeding factor from the data is about three. There is practically <u>no</u> probability that the seeding factor is less than one, or namely a negative effect. There is a small, but finite probability, that it exceeds five. Figures 13 b and 13c show the prior probability distributions of the gamma scale parameter β implied by the uniform prior probability of F in figures 13a and 13b. It is important to note that a diffuse prior on seeding factor is anything but diffuse for β and vice-versa. It would, therefore, seem preferable in the future to assign priors to F, wherever possible, rather than to functions of F for which there is less physical meaning.

5.2 Seeding Factor with the Transformed Rainfall Data

It appears desirable also to conduct a direct seeding factor analysis with the transformed data, since the distribution of these are nicer looking gamma functions with larger shape parameters. These more normal distributions might possess advantages for some types of investigation.

Here we define a transformed seeding factor F'

$$F' = \frac{\langle R' \rangle_{\text{seeded}}}{\langle R' \rangle_{\text{control}}} \simeq F^{0.25} \tag{37}$$

for use with the transformed data. (See section 3 for a discussion of the above approximation.) We consider only uniform prior probability on F'. Analogously with (24), (33) and (34) we proceed as follows:

$$p(F') = \frac{1}{b'-a'} \text{ for } a' \leq F' \leq b' \tag{38}$$

so that the posterior probability density distribution for F' is

$$p(F'|D') = C' \, p(D'|F') \quad \text{for } a' \leq F' \leq b'$$

$$= 0 \qquad \text{elsewhere}$$

$$= C' F'^{(-n\alpha')} \exp\left[- \frac{\alpha' \sum_{i=1}^{n} R'_i}{\langle R' \rangle_{NS} \, F'} \right] \quad \text{for } a' \leq F' \leq b'$$

$$= 0 \qquad \text{elsewhere} \tag{39}$$

where R' now stands for transformed or fourth root data and α' is seven.

We consider the two cases where F' is uniform in two ranges: first corresponding to F from 0.5 to 10 and second to F from 0.8 to 5, as previously.

The inverse transformations are

$$p(F) = \frac{1}{4(b-a)} F^{-0.75} \tag{40}$$

and

$$p(F|D) = \frac{C'}{4} F^{-\frac{(n\alpha+3)}{4}} \exp\left[\frac{-\alpha \sum_{i=1}^{n} R'_i}{\langle R'\rangle \quad F^{0.25}}\right] \tag{41}$$

The prior expectation for F is readily obtained by integration, while the posterior expectation can readily be shown (by transformation of the appropriate integral) to be

$$\langle F|D'\rangle = \langle F'^4|D'\rangle$$

$$= \left(\frac{\alpha' \sum_{i=1}^{n} R'_i}{\langle R'\rangle_{NS}}\right)^4 \frac{\gamma(n\alpha'-5,\mu/a) - \gamma(n\alpha'-5\ \mu/b)}{\gamma(n\alpha'-1,\mu/a) - \gamma(n\alpha'-1,\mu/b)} \tag{42}$$

or the fourth moment of the F' distribution.[8]

The results of this approach are illustrated graphically in figure 14. The slightly higher values of posterior expected seeding factors here than previously are attributed in part to the positive effect of the transform (fig. 5 and table 5) and in part to the difference influence of the prior probability.

---

[8] When normalized incomplete gamma functions are used in (42), the result must be multiplied by $1/[n\alpha'-2)\ (n\alpha'-3)\ (n\alpha'-4)\ (n\alpha'-5)$.

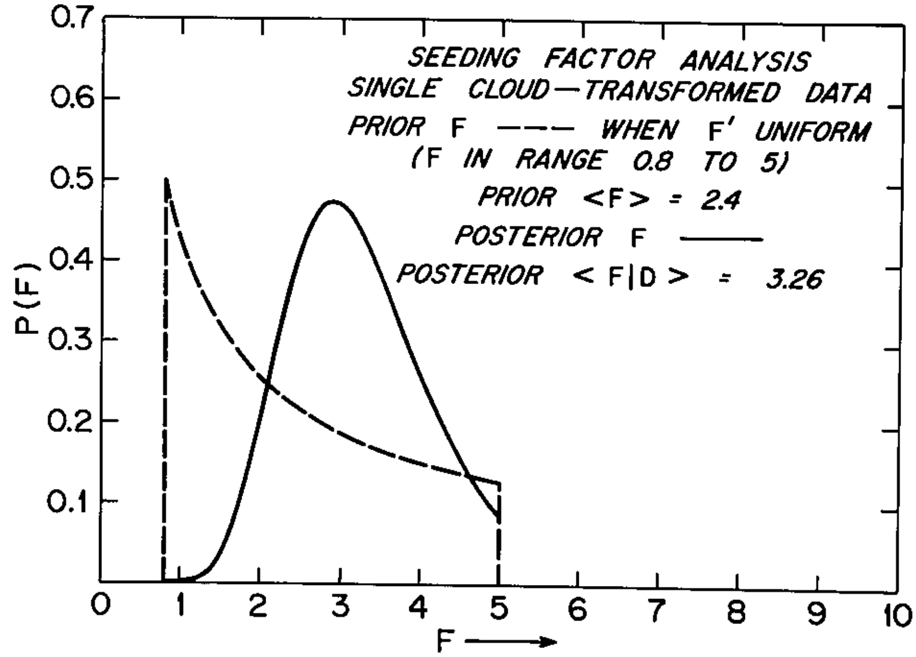

Figure 14. Direct seeding factor analysis. Transformed single cloud data. Prior probability on F' (fourth root of F) is uniform (dashed). Posterior probability on F is solid. a. F in range 0.5 to 10.

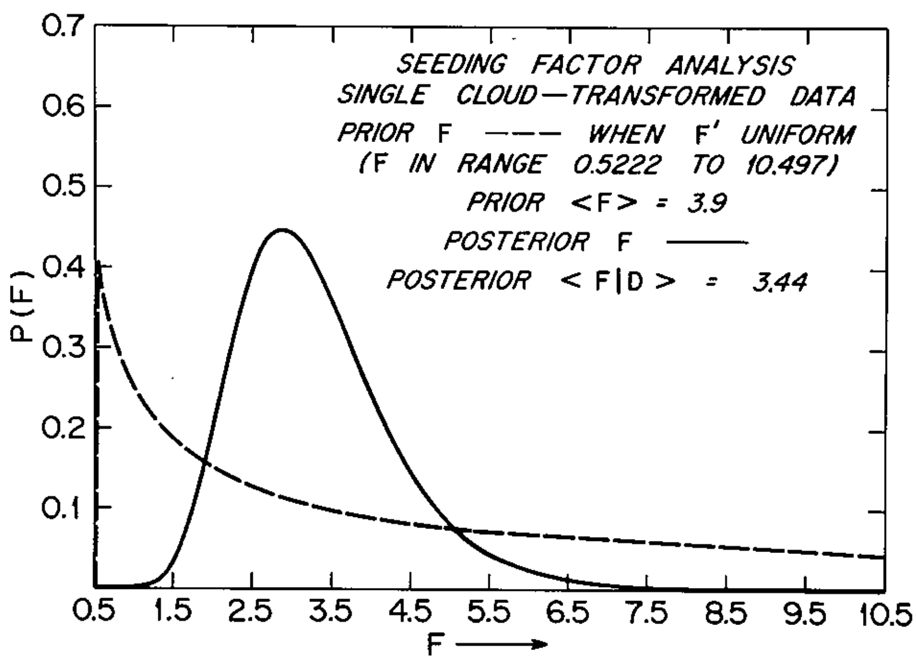

Figure 14b. F in range 0.8 to 5.

6. ON THE NUMBER OF OBSERVATIONS REQUIRED

An irremovable obstacle in most weather modification experiments is the small size of the data sample. In the 1968 EML single cloud experiment, 19 GO cases (14 seeded and five control) were not enough to separate the seeded and control populations to the five percent level by classical statistics, while increasing the sample to 52 GO cases in 1970 accomplished this goal. In approaching any experiment, we must inquire how many cases will be necessary to resolve various postulated magnitudes of seeding effect to specified degrees of significance and/or accuracy.

The demonstration that the rainfall observations are well fitted by gamma distributions provides a strong tool to attack this important problem. It is not difficult to generate on a computer any number of randomly chosen "rainfall" observations from any gamma distribution. This is cheaper than an observational program and much cheaper and less complicated than any actual modification program. We may then use simulated rainfall observations to address questions like the following:

1) How many observations are required to obtain the sample average of a rainfall population to a specified accuracy?
2) How many cases are needed to determine the shape and scale parameters of the distribution adequately? What is an adequate determination of these parameters for the particular experiments we are conducting?
3) What are the relative advantages and disadvantages in working with the raw or transformed data regarding the sampling

problem and distribution stability?

4) What are the magnitudes of errors in seeding factor that can arise from the data sampling problem and how many cases are required to ensure these do not exceed a specified level of acceptability?

Some aspects of these questions are answered in this section for the single cloud case, and a framework is established for their further pursuit, particularly in the area experiment.

The computer program used here, together with DAMAX, is called RAIN (listed in Appendix I). In the first part, it calculates integrated probability as function of βR, for any gamma function, given α. In the second part, it generates random numbers from 0 to 1 and interpolates in the integrated probability table, printing out R when β is given.

The simplest illustration involves sample averages. We take gamma distributions corresponding to the control rainfall populations, raw and transformed. Then we consider m samples of n observations, in order to assess our chances of determining the expected value $\langle R \rangle$ of the "real" gamma distribution within specified limits from a set of n observations. In preparing table 13, we started with m = 100. With parameters corresponding to the raw data, results were clearly not reproducible from one batch of 100 sample averages to the next. In order to obtain reproducible results, m was increased until this goal was achieved, namely to m = 1000. Appendix II shows from probability theory that results nearly identical to those of table 13 are obtained with the sample sizes given.

Table 13. Percent Probability That The Sample Mean of n Cases Lies Within Specified Limits of <R> - from Numerical Experiments.

A. Raw Data (m = 1000)

n	<5%	<10%	<20%	>30%	<1/2	>2
5	6.0	12.4	25.0	61.9	20.1	6.0
10	9.8	19.5	37.7	45.9	7.1	2.1
20	13.9	28.6	51.2	28.1	1.8	0
50	20.5	39.1	72.3	7.4	0	0

B. Transformed (m = 1000)

n	<5%	<10%	<20%	>30%
5	21.3	45.0	78.1	5.6
10	33.1	55.2	90.2	1.0
20	43.6	76.6	98.5	0
50	68.0	94.0	100	0

With the raw data, note that doubling the sample size leads to roughly a 50 percent increase in the percentage of cases falling within the 5, 10 and 20 percent categories. With a sample of ten cases, there is still a serious possibility of getting a sample average in error by a factor of two, which virtually disappears when we obtain 20 cases.

In considering the transformed data, there is little, if any, indication of reproducibility occurring sooner than with the raw data. With the transformed data, it should be recalled that 5 percent error corresponds to roughly 20 percent error in the raw data and hence from this viewpoint there appears to be no advantage in working with transformed data.

An important further test with this program is to simulate a "seeding" experiment analogous to the manner of Huff (1971) in which one-half of the randomly generated sample averages from the same population[9] are arbitrarily called "seeded." The ratio of each term in two matrices containing 100 "seeded" and 100 "control" samples was taken and results are shown in table 14. This table presents the frequency of various ratios of the "seeded" sample average to the "control" sample average. This kind of spurious result could arise from an attempted modification experiment where the treatment had no effect.

Table 14. Simulated "Seeding" Experiment - 100 Samples of n Cases each. "Seeded" to "Control" Ratio.

	Frequency Distribution of Ratios			
n=	5	10	20	50
min.	.058	0.26	0.24	.56
≤ 0.50	17	7	8	0
0.51 - 0.80	24	27	19	21
0.81 - 1.20	22	23	40	53
1.21 - 1.50	7	15	16	21
1.51 - 2.0	7	11	12	5
2.01 - 2.50	10	4	3	0
2.51 - 3.0	0	10	2	0
≥ 3.01	13	3	0	0
max.	10.70	4.06	2.71	1.81

Table 14 suggests extreme caution in drawing inferences from seeding experiments with small samples of data. It also suggests that to specify natural distributions adequately for resolution of seeding effects of a factor of two to three, 20 to 50 cases are a necessary minimum. If we come down to dealing with expected seeding factors of 50 percent or less, it is plain that 50 cases are not adequate, with this type of data distribution.

The final important question here is the degree of accuracy

[9] That is, all cases are taken from the same "real" gamma distribution.

with which the gamma function parameters can be recaptured as a function of n, the number of observations. Here we consider two gamma functions, with shape and scale parameters chosen to correspond to the control single cloud data, raw and transformed, respectively. We select n observations at random from these distributions, pretend the set is a set of rainfall data for n clouds and apply DAMAX. We repeat this procedure m times for each value of n and then examine the statistics of the recovered parameters and their departures from the "real" parameters. In tables 15 and 16 to follow we take n = 20, 50 and 100 successively; the reason for the choice is that these are the numbers of control cases we might expect to obtain in single or multiple cumulus experiments in two to ten years of work.

Here we take m = 100 (100 sets of five, 20 and 50 cases).[10] Expert opinions consulted suggest that m = 100 may be marginal for reproducibility, but in analyzing the latter we are aided by the extensive valuable work of Bowman and Shenton (1968; 1970) on the gamma distribution. They derived asymptotic expansions of the expectations, presenting tables of bias, standard deviations, etc. of the parameters as a function of the parameters and of sample size. They also conducted Monte Carlo experiments with up to 10^5 cases each.

Table 15 presents results for the gamma function corresponding to raw single cloud data, while table 16 corresponds to the transformed single cloud data.

[10] The limit to the sample at this point is imposed by the cost of computer time.

Table 15. Application of DAMAX to "Data" Generated by Random Selection from Specified Gamma Functions.

Gamma function corresponding to raw control data
$\alpha = 0.6$ $\beta = 0.00364$ $\langle R \rangle = 164.5885$ acre-ft

A. n = 20				
	α	$\beta \times 10^2$	Prob. Gamma	$\overline{R}$ (acre-ft)
Max. value	1.1092	1.1870	0.4888	327.1277
Min. value	0.3844	0.1730	0.0441	73.9079
Mean	0.6677	0.4769	0.2689	154.4864
Variance	0.0275	0.0392	0.0058	2384.9482
Std. dev.	0.1658	0.1979	0.0762	48.8359
B. n = 50				
Max. value	0.9947	0.6910	0.6408	265.3409
Min. value	0.3959	0.2120	0.1034	91.8385
Mean	0.6137	0.3779	0.3679	168.6239
Variance	0.0124	0.0116	0.0121	880.1806
Std. Dev.	0.1115	0.1075	0.1102	29.6678
C. n = 100				
Max. value	0.8516	0.5620	0.8166	218.2472
Min. value	0.4659	0.2500	0.0762	118.1907
Mean	0.6026	0.3672	0.4596	167.2139
Variance	0.0055	0.0052	0.0334	498.9574
Std. Dev.	0.0743	0.0722	0.1827	22.3374

Table 16. Application of DAMAX to "Data" Generated by Random Selection from Specified Gamma Functions.

Gamma function corresponding to transformed control data

$\alpha = 7.0$ $\beta = 2.22360$ $\langle R \rangle = 2.93353\ (\text{acre-ft})^{0.25}$

	α	β	Prob. Gamma	$\overline{R}$ (acre-ft)$^{0.25}$
A. n = 20				
Max. value	29.6933	9.9008	0.2449	3.4520
Min. value	2.7792	1.0132	0.0133	2.2276
Mean	8.1524	2.8038	0.1684	2.9308
Variance	12.3459	1.5181	0.0024	0.0468
Std. dev.	3.5137	1.2321	0.0492	0.2164
B. n = 50				
Max. value	12.3816	4.1242	0.5065	4.1160
Min. value	4.4934	1.4470	0.0306	2.4727
Mean	7.2542	2.4750	0.2596	2.9503
Variance	2.3753	0.2552	0.0122	0.0459
Std. dev.	1.5412	0.5052	0.1103	0.2143
C. n = 100				
Max. value	10.6011	3.5235	0.7152	3.2217
Min. value	5.2579	1.7100	0.0072	2.6773
Mean	7.3754	2.5223	0.3972	2.9302
Variance	1.1397	0.1492	0.0298	0.0113
Std. dev.	1.0676	0.3863	0.1726	0.1064

It is noteworthy that, despite the fact that all observation sets were selected from "real" gamma distributions, the gamma distribution did not always come out the most probable, with the maximum entropy criterion. Its rank is tabulated in table 17.

Table 17. Rank of Gamma Distribution.

	A. Raw Data					B. Transformed Data				
n/rank	1	2	3	4	5	1	2	3	4	5
20	14	65	14	5	2	17	17	18	48	0
50	41	54	5	0	0	34	30	12	24	0
100	58	42	0	0	0	54	26	14	5	0

Next we examine the bias of the mean and the variance that we might expect from a much larger number of calculations. Table 18 is reproduced from the figures of Bowman and Shenton (1970).

Table 18. (After Bowman and Shenton, 1970). Properties of Recaptured Gamma Parameters as Function of Sample Size.

	Mean Bias		Fractional Variance	
	α	$1/\beta$	α	$1/\beta$
A. α = 0.6 (corresponding to raw single cloud data)				
n				
20	0.1311	-0.0425	0.1204	0.1503
50	0.0477	-0.0169	0.0344	0.0610
100	0.0231	-0.0084	0.0155	0.0306
B. α = 7 (corresponding to transformed single cloud data)				
n				
20	0.1710	-0.0499	0.1757	0.0977
50	0.0619	-0.0200	0.0480	0.0403
100	0.0300	-0.0100	0.0213	0.0203

Our small samples in tables 15 and 16 were extensively compared with the results in table 18. There is fair agreement in bias, improving as n increases, and excellent agreement in variance. In no cases was our bias significantly or systematically larger than the above. These results increase our confidence in our RAIN program and random number generator, which was subjected to numerous independent tests. In all but the variance of the scale parameter, the advantage is with the smaller shape parameter; hence table 18 provides no reason for preference to work with transformed rather than raw data.

It appears from tables 15 - 18 that essential improvement in parameter recovery is secured by progressing from 20 to 50 cases, but less is gained in doubling that sample from 50 to 100. However, further consideration of the variances and standard deviations presented versus anticipated seeding factors presses the argument for 100 cases.

If a seeding factor is two or more, it is likely that about 50 pairs of cases can resolve it adequately. If we deal with seeding factors of much less than 2, then we must attempt to obtain roughly 100 cases. Furthermore, with seeding factors less than about 2, radar calibration errors and/or gaging problems, not considered in this paper, become a serious consideration.

7. CONCLUDING REMARKS

The Bayesian approach applied to the single cumulus experiments confirms a seeding factor of about 3 on the rainfall. This result can be shown relatively independent of widely differing prior probability assumptions.

The approach leads to valuable numerical experiments on the number of cases needed to resolve various sizes of seeding effects. The entire framework will next be applied to the multiple c ulus experiments of 1970, 1971 and 1972.

8. ACKNOWLEDGMENTS

The writers gratefully acknowledge the inspiration and also detailed help by Dr. Myron Tribus whose encouragement and example provided the impetus for this work.

Many NOAA colleagues have been extremely helpful. Dr. William L. Woodley has provided continuous assistance. Glenn Brier and Gerald F. Cotton have contributed effectively and patiently on the statistical aspects. Mr. H. C. S. Thom has offered valuable expertise on gamma functions, as has Dr. Albion Taylor. Outside of NOAA, Professor C. T. Haan of the University of Kentucky and Professor L. R. Shenton of the University of Georgia provided useful help on several calculations. Stimulating discussions of the Bayesian aspects of the problem were held with Professor John Pratt and Dr. Bing Sung of the Harvard Business School.

Within EML, we wish to thank Robert N. Powell for excellent drafting work, and Mrs. Constance Arnhols for a good job in preparing the manuscript. We also appreciate the fine drafting help by Mr. Robert Carrodus of the NOAA National Hurricane Research Laboratory.

9. REFERENCES

Barger, G. L., R. H. Shaw and R. F. Dale (1959), Gamma distribution parameters from 2- and 3-week precipitation totals in the North Central Region of the United States. Agricultural and Home Economics Experiment Station, Iowa State University, Ames, Iowa, 183 pp.

Bowman, K. O. and L. R. Shenton (1968), Properties of estimators for the gamma distribution. CTC-1. Computing Technology Center, Union Carbide Corporation, Nuclear Division, Oak Ridge, Tennessee, 50 pp.

Bowman, K. O. and L. R. Shenton (1970), Small sample properties of estimators for the gamma distribution. CTC-28. Computing Technology Center, Union Carbide Corporation, Nuclear Division, Oak Ridge, Tennessee, 160 pp.

Bridges, T. C. and C. T. Haan (1972), Reliability of precipitation probabilities estimated from the gamma distribution, Monthly Wea. Rev. 100, 607-611.

Herndon, A., C. L. Courtright, W. L. Woodley and J. J. Fernandez-Partagas (1971), A south Florida radar raingage comparison for 1970, NOAA Tech. Memo. ERL OD-7, 22 pp.

Huff, F. A. (1971), Evaluation of precipitation records in weather modification experiments. Advances in Geophysics 15, 59-135.

Kendall, M. G. and A. Stuart (1963), The Advanced Theory of Statistics, (2nd edition), 1 , Distribution Theory, The Hafner Publishing Co., N.Y., N.Y., 431 pp.

Pearson, K. (Editor) (1951), Tables of the Incomplete Gamma Function, Cambridge University Press, Cambridge, England, 164 pp.

Mooley, D. A. and H. L. Crutcher (1968), An application of gamma distribution function to Indian rainfall, ESSA Tech. Rep. EDS 5, Environmental Data Service, Silver Spring, Maryland, 47 pp.

Mooley, D. A. (1972), An estimate of the distribution and stability period of the parameters of the gamma probability model applied to monthly rainfall over Southeast Asia during the summer monsoon. Monthly Wea. Rev., in press.

Shenton, L. R. and K. O. Bowman (1970), Remarks on Thom's estimators for the gamma distribution, Monthly Wea. Rev. 98, 154-160.

Simpson, J. (1972), Use of the gamma distribution in single cloud rainfall analysis, Monthly Wea. Rev. 100, 309-312.

Simpson, J. and W. L. Woodley (1971), Seeding cumulus in Florida: new 1970 results, Science 9, 117-126.

Simpson, J. and J. Pezier (1971), Outline of a Bayesian approach to the EML multiple cloud seeding experiments, NOAA Tech. Memo ERL OD-8, 43 pp.

Simpson, J., W. L. Woodley, A. H. Miller and G. F. Cotton (1971), Precipitation results of two randomized pyrotechnic cumulus seeding experiments, J. Appl. Meteor. 10, 526-544.

Thom, H. C. S. (1947), A note on the gamma distribution, manuscript, Statistical Laboratory, Iowa State College, Ames, Iowa.

Thom, H. C. S. (1957), A statistical method of evaluating augmentation of precipitation by cloud seeding, Tech. Report No. 1, Advisory Committee on Weather Control, Washington, D.C., 62 pp.

Thom, H. C. S. (1958), A note on the gamma distribution, Monthly Wea. Rev. 86, 117-122.

Thom, H. C. S. (1968), Direct and inverse tables of the gamma distribution, Tech. Report EDS-2, Environmental Science Services Administration, Environmental Data Service, Silver Spring, Maryland, 30 pp.

Thom, H. C. S. and I. B. Vestal (1968), Quantiles of monthly precipitation for selected stations in the contiguous United States, Tech. Report EDS-6, Environmental Science Services Administration, Environmental Data Service, Silver Springs, Maryland, 5 pp. and tables.

Tribus, M. (1969), Rational descriptions, decisions and designs, Pergamon Press, N.Y., N.Y., 478 pp.

Woodley, W. L. (1970), Precipitation results from a pyrotechinic cumulus seeding experiment, J. Appl. Meteor. 9, 242-257.

Woodley, W. L. and A. Herndon (1970), A raingage evaluation of the Miami reflectivity-rainfall rate relation, J. Appl. Meteor. 9 , 258-264.

APPENDIX I

LISTING OF PROGRAM RAIN

```
      DIMENSION P(100),R(1000),RV(1000)
      DIMENSION PAK(6),BUFF(14),ELTNM(2)
      INTEGER DATE(2)
      ABNORMAL USRRND
      DATA PAK/'RANDOM','RAWDAT','      ','      ','      ','      '/
  200 READ (5,16) E1,N1,A,Q,B,N,M
   16 FORMAT (A6,I6,2F5.1,F8.5,2I4)
      IF (N.EQ.0) CALL EXIT
      DO 300 KL=1,100
      N1=N1+1
      ENCODE (12,15,ELTNM,IX)E1,N1
   15 FORMAT (A6,J6)
      PAK(3)=ELTNM(1)
      PAK(4)=ELTNM(2)
      RVSUM=0.
      CALL ERTRAN (9,DATE(1),DATE(2))
      DECODE (12,10,DATE,IX) IA,IB
   10 FORMAT (2I6)
      IP=(IA+IB)/615+IB
      S=USRRND(IP)
      DO 20 I=1,50
      Y=A*I/50
      P(I)=FNI(A,Y)
   20 CONTINUE
      DO 25 I=51,100
      Y=A+(Q-A)*(I-50)/50
      P(I)=FNI(A,Y)
   25 CONTINUE
      DO 60 K=1,M
      RSUM=0.
      DO 50 J=1,N
      S=USRRND(0)
      DO 30 I=1,100
      IF (S.LT.P(I)) GO TO 35
   30 CONTINUE
   35 IF (I.GT.50) GO TO 40
      IF (I.EQ.1) GO TO 45
      R(J)=A*(I-1)/50+(A/50)*(S-P(I-1))/(P(I)-P(I-1)))
      X0=A*(I-1)/50
      X1=A*I/50
      PX0=P(I-1)
      PX1=P(I)
      GO TO 48
   40 RR=A+(Q-A)*(I-51)/50
      R(J)=(RR+((Q-A)/50)*(S-P(I-1))/(P(I)-P(I-1)))
      X0=A+((Q-A)*(I-51)/50)
```

LISTING OF PROGRAM RAIN (Continued)

```
      X1=X0+((Q-A)/50)
      PX0=P(I-1)
      PX1=P(I)
      GO TO 48
   45 R(J)=A/50*S/P(I)
      X0=0.0
      X1=A/50
      PX0=0.0
      PX1+P(I)
   48 X=R(J)
   99 PX=FNI(A,X)
      IF ((ABS(PX-S)).LE.1E-4)GO TO 101
      IF (PX-S) 100,101,102
  100 X0=X
      PX0=PX
      X=X0+((S-PX0)/(PX1-PX0))*(X1-X0)
      GO TO 99
  102 X1=X
      PX1=PX
      X=X0+((S-PX0)/(PX1-PX0))/(X1-X0)
      GO TO 99
  101 R(J)=X/B
      RSUM=RSUM+R(J)
   50 CONTINUE
      RV(K)=RSUM/N
      RVSUM=RVSUM+RV(K)
   60 CONTINUE
      AVE=RVSUM/M
      WRITE (6,55) (RV(K),K=1,M)
   55 FORMAT (1H,8F10.5)
      WRITE (6,65) AVE
   65 FORMAT (1H,'AVERAGE =',F10.5)
      CALL USRSDO(PAK)
      ENCODE (80,78,BUFF,ITRN) M
      CALL USRSDW(BUFF)
      DO 75 IV=1,M,8
      IK=IV+7
      ENCODE (80,77,BUFF,ITRN) (RV(L),L=IV,IK)
   77 FORMAT (8F10.5)
   78 FORMAT (I10)
   75 CALL USRSDW(BUFF)
      CALL USRSDC
  300 CONTINUE
      GO TO 200
      END
```

LISTING OF PROGRAM RAIN (Continued)

```
      FUNCTION FNI(X,Y)
      IF (Y.LT.(X/2+4)) GO TO 120
      G7=1.
      G8=1.
      L=IDINT(X)
      DO 100 I=1,L
      G7=G7*(X-I)/Y
      IF (G7.LT.1E-4) GO TO 110
      G8=G8+G7
 100  CONTINUE
 110  FNI=1.-G8*EXP((X-1.)*LOG(Y)-Y-FNG(X))
      GO TO 150
 120  G8=1./X
      G7=G8
      DO 130 J=1,50
      G7=G7*Y/(X+J)
      IF ((G7*X).LT.1E-4) GO TO 140
      G8=G8+G7
 130  CONTINUE
 140  FNI=G8*EXP(X*LOG(Y)-Y-FNG(X))
 150  RETURN
      END

      FUNCTION FNG(R)
      IF (R.LT.4.) GO TO 3260
      G=R*(DLOG(R)-1.)+0.5*DLOG(6.2831853/R)
      FNG=G+(1-1/(30*R*R))/(12*R)
      GO TO 3350
3260  G1=R-IDINT(R)
      LG2=IDINT(R)-1
3280  G = 1.-(.57710166-(.98585399-(.87642182-(.8328212-(.5684729
     * -(.25482049-.0514993*G1)*G1)*G1)*G1)*G1)*G1)*G1
      IF (LG2) 3310,3340, 3320
3310  G = G/R
      GO TO 3340
3320  CONTINUE
      DO 3330 LG9=1,LG2
      G=G*(G1+LG9)
3330  CONTINUE
3340  FNG=DLOG(G)
3350  RETURN
      END
```

APPENDIX II

DISTRIBUTIONS AND PROPERTIES OF $\bar{R}$ AND $\hat{F}$

by Anthony Olsen

In section 6 consideration was given to the number of observations required to accomplish specific goals. The methodology used consisted of assuming that the rainfall observations arose from a gamma distribution with known parameters. This assumption is supported in the earlier sections. Utilizing the gamma distribution, rainfall data were simulated and Monte Carlo procedures were used to answer questions concerning sample size, seeding factor and other variables. The purpose of this appendix is to give an alternate approach for answering some of the questions. Specifically, the number of observations required to obtain the sample average of a rainfall population to a specified accuracy and the natural variabilities of the sample seeding factor is a function of the sample size and true seeding factor.

The basic idea of the method utilized is as follows. It is assumed that the rainfall observations are observed values of a random variable that has a gamma distribution. From this distributional assumption, the sample distributions of the sample mean $\bar{R}$ and the sample seeding factor $\hat{F}$ are derived. It is then a simple matter to construct the appropriate probability statements in answer to questions of natural variability. The difference between the present approach and that of section 6 is the manner in which the sample distributions are obtained. The latter uses a simulation approach. Each is equally valid.

Let R_1, R_2, ..., R_n be random samples from a gamma distribution with parameter α and β. By the reproductive property of the gamma distribution it follows that the sample mean

$$\overline{R} = \frac{1}{n} \sum_{i=1}^{n} R_i$$

has a gamma distribution with parameters $n\alpha$ and $n\beta$. Moreover, the expected value of $\overline{R}$ is α/β, the same as the original population.

The question of interest is: what is the percent probability that the sample mean of n observations lies within the specified limits of <R>? In terms of a probability statement

$$Pr[|\overline{R} - \langle R \rangle| \leq b \langle R \rangle]$$

gives the probability that $\overline{R}$ is within 100b percent of <R>. Statements of the above form only require the use of incomplete gamma tables to determine the probability, since $\overline{R}$ has a gamma distribution. Note that the distribution changes parameters as n changes.

Table I corresponds to table 13 given in section 6 with the probabilities calculated using table 7 in Biometrika Tables for Statisticians by E. S. Pearson and H. O. Hartley (1966), the normal approximation to the gamma distribution given by (26.4.14). Abramowitz and Segun (1964) was utilized for values outside the range of the table.

In comparing the simulated probabilities with the probabilities given here, two items are of interest. First, note the close agreement between the tables. Second, note that table 13 gives an approximation to the actual percent probabilities presented in table I.

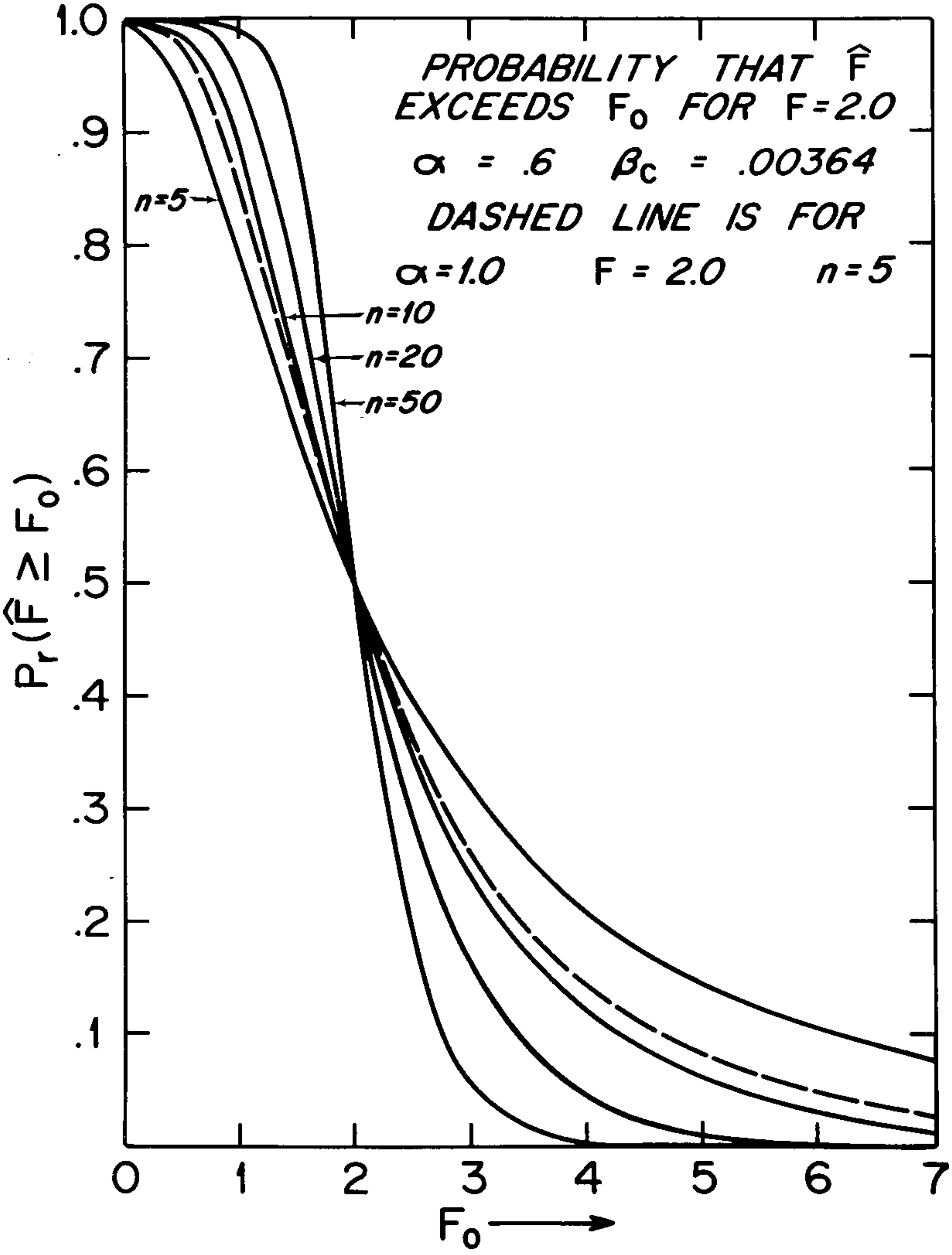

Figure 4

☆ U.S. GOVERNMENT PRINTING OFFICE: 1973 — 784-213 / 1142 REGION NO. 8

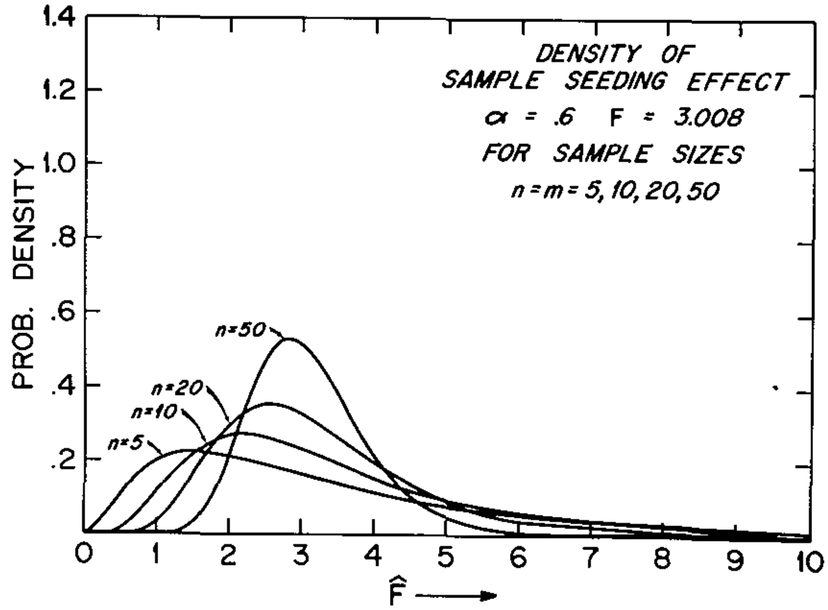
DENSITY OF
SAMPLE SEEDING EFFECT
α = .6 F = 3.008
FOR SAMPLE SIZES
n = m = 5, 10, 20, 50
n=50
n=20
n=10
n=5
PROB. DENSITY
1.4
1.2
1.0
.8
.6
.4
.2
0
1
2
3
4
5
6
7
8
9
10
F̂

Figure 3

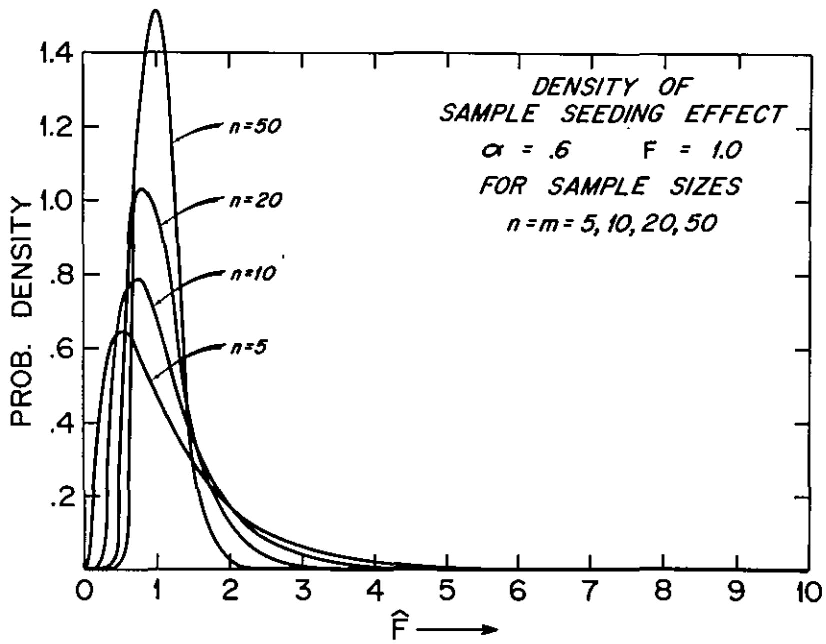

Figure 1

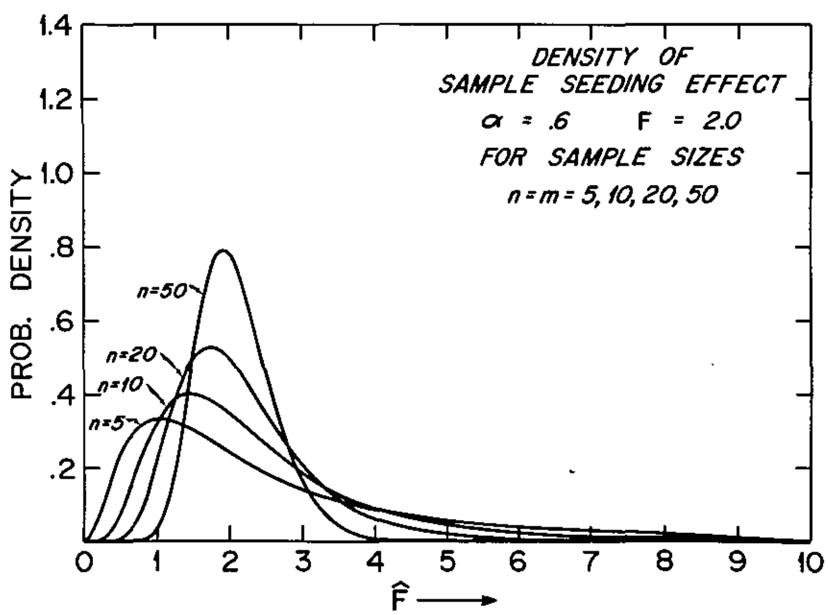

Figure 2

Table VI presents the same situation as table V with the exception of the change in α. The effect of α is readily seen to be an apparent increase in sample when α is increased. This is as expected theoretically. The sampling distribution of $\hat{F}$ depends on α only in the form $m\alpha$ or $n\alpha$. Hence, an increase of α increases the product. Furthermore, in calculating probability statements $\nu_1 = 2n\alpha$ and $\nu_2 = 2m\alpha$ enter as the degrees of freedom associated with Snedecor's F-distribution.

In the actual application of the distribution of sample seeding factor, the value of α is not known, but is estimated from both the seeded and control rainfall. If α is overestimated, the effect is that the probability statements made will be overly optimistic. For example, using tables V and VI we obtain

$$\Pr\left[1.5 \le \hat{F} \le 2.5 \,\middle|\, \begin{matrix} n = 5 \\ \alpha = .6 \end{matrix}\right] = .23539$$

$$\Pr\left[1.5 \le \hat{F} \le 2.5 \,\middle|\, \begin{matrix} n = 5 \\ \alpha = 1 \end{matrix}\right] = .30548.$$

The change in α from 0.6 to 1.0 has the effect of increasing the degrees of freedom associated with Snedecor's F-distribution from $\nu_1 = \nu_2 = 6$ to $\nu_1 = \nu_2 = 10$ for $m = n = 5$. Therefore, in the interpretation of the probabilities statements some care must be exercised.

REFERENCES

Abramowitz, M. and I.A. Segun (Editors) (1964), Handbook of Mathematical Functions, National Bureau of Standards, 1046 pp.

Pearson, E.S. and H.O. Hartley (Editors) (1966), Biometrika Tables for Statisticians, Cambridge University Press, Cambridge, England.

Pearson, K. (Editor) (1968), Tables of the Incomplete Beta-Function, Cambridge, England, 505 pp.

The effect of changing sample size on the probability distribution of $\hat{F}$ is apparent from tables I - V. As the sample size increases, the variance of the distribution decreases. This pattern is present for the three different true seeding factors considered. Instead of graphing the probabilities, it is informative to graph the probability density functions. Figures 1 to 3 illustrate the results. The decrease in dispersion as the sample size increases is again illustrated. Also, for a fixed sample size, the dispersion increases with increased seeding factor. This is most notable when comparing the curves for (n = 50 and F = 1.0) and (n = 50 and F = 3.008). This is expected since the variance of $\hat{F}$ is

$$F^2\left(\frac{m\alpha}{m\alpha-1}\right)^2\frac{n\alpha + m\alpha-1}{n\alpha(m\alpha-2)} .$$

Table A11-VI. Percent Probability That The Sample Seeding Factor Will Exceed F_o When The True Seeding Factor Is 2.0 (i.e., $\alpha = 1.0$, $\beta_c = .00364$, $\beta_s = .00182$).

F_o	Sample Size n: 5	10	20	50
0.0	100.00	100.00	100.00	100.00
0.5	98.042	99.842	99.999	100.00
0.8	91.763	97.657	99.763	100.00
1.2	78.328	86.885	94.447	99.414
1.5	67.103	73.691	81.635	92.347
2.0	50.000	50.000	50.000	50.000
2.5	36.555	31.141	24.220	13.427
3.0	26.657	18.609	10.206	2.193
4.0	14.494	6.495	1.567	0.033
5.0	8.236	2.342	0.237	0.000
6.0	4.893	0.890	0.037	0.000
7.0	3.044	0.367	0.007	0.000
8.0	1.958	0.158	0.001	0.000
9.0	1.309	0.074	0.000	0.000
10.0	0.900	0.036	0.000	0.000

Table AII-IV. Percent Probability That The Sample Seeding Factor Will Exceed F_0 When The True Seeding Factor is 3.008 (i.e., α = .6, β_σ = .00121, β_c = .00364).

F_o	Sample Size n: 5	10	20	50
0.0	100.00	100.00	100.00	100.00
0.5	97.68	99.79	99.99	100.00
0.8	93.41	98.51	99.90	100.00
1.2	85.59	93.73	98.56	99.97
1.5	79.10	87.87	95.24	99.60
2.0	68.37	75.49	83.79	94.15
2.5	58.60	62.31	67.28	76.15
3.0	50.13	50.18	50.26	50.42
4.0	36.91	31.47	24.53	13.65
5.0	27.63	19.57	11.05	2.60
6.0	21.08	12.31	4.89	0.43
7.0	16.38	7.89	2.18	0.07
8.0	12.96	5.18	1.01	0.01
9.0	10.41	3.47	0.47	0.00
10.0	8.48	2.38	0.23	0.00

Table A11-V. Percent Probability That The Sample Seeding Factor Will Exceed F_o When The True Seeding Factor Is 2.0 (i.e., α = .6, β_c = .00364, β_s = .00182).

F_o	Sample Size n: 5	10	20	50
0.0	100.00	100.00	100.000	100.000
0.5	94.208	98.835	99.940	100.000
0.8	85.525	93.673	98.542	99.973
1.2	72.475	80.553	89.070	97.462
1.5	63.209	68.688	75.670	86.562
2.0	50.000	50.000	50.000	50.000
2.5	39.670	35.262	29.453	19.563
3.0	31.744	24.650	16.364	5.955
4.0	20.992	12.221	4.824	0.416
5.0	14.476	6.326	1.458	0.027
6.0	10.351	3.433	0.465	0.002
7.0	7.646	1.964	0.163	0.000
8.0	5.792	1.165	0.060	0.000
9.0	4.494	0.724	0.024	0.000
10.0	3.556	0.465	0.010	0.000

Table A11-II. Percent Probability That The Sample Seeding Factor Lies Within the Specified Intervals As A Function of Sample Size for The Raw Data (i.e., $\alpha = .6$, $\beta_s = \beta_c = .00364$).

n	5	10	20	50
0 - 0.5	20.993	12.216	4.821	0.415
0.5 - 0.8	18.779	23.046	24.632	19.131
0.8 - 1.2	18.619	26.881	37.609	56.301
1.2 - 1.5	9.865	13.207	16.574	18.198
1.5 - 2.0	10.748	12.434	11.543	5.540
2.0 - 2.5	6.521	5.891	3.364	0.388
2.5 - 3.0	4.123	2.892	0.992	0.025
> 3.0	10.352	3.433	0.465	0.002
Mean	1.50	1.20	1.09	1.034
Variance	3.75	0.66	0.228	0.075
Std. Deviation	1.936	0.812	0.478	0.274

Table A11-III. Percent Probability That The Sample Seeding Factor Lies Within Specified Intervals When It Is Assumed That The True Seeding Factor is 3.008 (i.e., $\alpha = .6$, $\beta_s = .00121$, $\beta_c = .00364$).

	Sample Size n			
Interval	5	10	20	50
0.0 - 0.5	2.317	0.205	0.003	0.000
0.5 - 0.8	4.278	1.283	0.093	0.000
0.8 - 1.2	7.813	4.785	1.340	0.026
1.2 - 1.5	6.495	5.857	3.323	0.377
1.5 - 2.0	10.728	12.378	11.450	5.443
2.0 - 2.5	9.771	13.186	16.507	18.002
2.5 - 3.0	8.471	12.123	17.022	25.736
3.0 - 4.0	13.218	18.712	25.733	36.769
4.0 - 5.0	9.279	11.899	13.477	11.048
5.0 - 6.0	6.548	7.259	6.162	2.177
6.0 - 7.0	4.698	4.423	2.710	0.362
7.0 - 8.0	3.422	2.709	1.173	0.055
8.0 - 9.0	2.556	1.714	0.533	0.010
9.0 - 10.0	1.930	1.091	0.241	0.001
> 10.0	8.476	2.376	0.233	0.000
Mean	4.512	3.610	3.282	3.112
Variance	40.299	5.967	2.068	0.679
Std. Deviation	5.824	2.443	1.438	0.824

$F = 3.008$ and

m	$E(\hat{F})$
20	3.282
50	3.112

An unbiased estimator for the true seeding factor is possible if α is assumed known. Define

$$\tilde{F} = \frac{m\alpha - 1}{m\alpha}\hat{F}$$

Then $E(\tilde{F}) = \frac{m\alpha-1}{m\alpha}E(\hat{F}) = \frac{\beta_s}{\beta_c} = F$. Usually, however, α must be estimated. It is possible to substitute in an estimate of α in determining $\tilde{F}$.

A table analogous to table 14 in section 6 may be constructed from the distribution of $\hat{F}$. The procedure followed is the same as in table I, the evaluation of probability statements of the form

$$Pr[a \le \hat{F} \le b \mid n, m, \alpha, \beta_s, \beta_c].$$

It is necessary to transform the probability statement into terms of a beta random variable and then make use of a table of the incomplete beta distribution. The necessary transformation is

$$B = \frac{1}{1 + \frac{\nu_1}{\nu_2}\frac{\beta_s}{\beta_c}\hat{F}}$$

so that

$$Pr[a \le \hat{F} \le b \mid n, m, \alpha, \beta_s, \beta_c]$$

$$= Pr\left[\left(1 + \frac{n\beta_s}{m\beta_c} b\right)^{-1} \le B \le \left(1 + \frac{n\beta_s}{m\beta_c} a\right)^{-1} \mid n, m, \alpha, \beta_s, \beta_c\right]$$

where B has a beta distribution with parameters $m\alpha$ and $n\alpha$.

the sample seeding factor is defined as $\hat{F} = \overline{R}_S/\overline{R}_C$. The sampling distribution of $\hat{F}$ is easily determined by recognizing that $\overline{R}_S$ and $\overline{R}_C$ are independent gamma distributions with parameters $n\alpha$, $n\beta_S$ and $m\alpha$, $m\beta_C$ respectively. Hence, except for a constant, depending upon the parameters, $\hat{F}$ has essentially Snedecor's F-distribution. That is, since $2n\beta_S\overline{R}_S$ and $2m\beta_C\overline{R}_C$ have chi-square distributions, it follows that

$$\frac{2n\beta_S\overline{R}_S/(2n\alpha)}{2m\beta_C\overline{R}_C/(2m\alpha)} = \frac{\beta_S}{\beta_C}\hat{F}$$

has Snedecor's F-distribution with $\nu_1 = 2n\alpha$ and $\nu_2 = 2m\alpha$ degrees of freedom. Hence, by specifying the parameters of the gamma distributions and the sample sizes, the sampling distribution of $\hat{F}$ is known to be β_C/β_S times a Snedecor's F-distribution.

The mean and variance of $\hat{F}_s$ are determined from the mean and variance of Snedecor's F-distribution

$$\frac{m\alpha}{m\alpha-1} \quad \text{and} \left(\frac{m\alpha}{m\alpha-1}\right)^2 \frac{n\alpha + m\alpha-1}{n\alpha\ (m\alpha-2)}$$

for $m\alpha>1$ and $m\alpha>2$ respectively.

Hence $\langle\hat{F}\rangle = E(\hat{F}) = \frac{\beta_C}{\beta_S}\frac{m\alpha}{m\alpha-1}$

and
$$\mathrm{Var}(\hat{F}) = \frac{\beta^2{}_C}{\beta^2{}_S}\left(\frac{m\alpha}{m\alpha-1}\right)^2 \frac{n\alpha+m\alpha-1}{n\alpha(m\alpha-2)}$$

for $m\alpha>1$ and $m\alpha>2$ respectively. In general, $E(\hat{F})>F$ so that on the average if F is the estimator for seeding factor, then the actual seeding factor is being overestimated. As the control sample size increases, the bias decreases. For example, let $\alpha = .6$, $\beta_S = .00121$ and $\beta_C = .00364$, then

Table A11-I. Percent Probability That The Sample Mean of n Cases Lies Within Specified Limits of <R>. (Probabilities Obtained From Gamma Tables.)

A. Raw data (α = .6, β = .00364)

n	<5%	<10%	<20%	>30%	<1/2	>2
5	8.218	13.420	26.696	60.350	19.115	6.197
10	9.615	19.128	37.493	45.728	8.392	2.034
20	13.645	26.937	51.302	29.154	2.009	0.252
50	21.456	41.617	72.999	9.654	0.042	0.001

B. Transformed data (α = 7, β = 2.38620)

n	<5%	<10%	<20%	>30%
5	23.233	44.636	76.666	7.324
10	32.425	59.805	90.732	1.266
20	44.600	77.090	98.192	0.057
50	65.064	93.885	99.976	0.000

The second major investigation concerns the natural variability of the sample seeding factor $\hat{F}$. In section 6 a simulated "seeding" experimen was performed using the assumption that the rainfall observations were a random sample from a gamma distribution. It was assumed that the true seeding factor was one, i.e. that the control and seeded observations were from the same distribution. In the following development the sampling distribution of $\hat{F}$ will be determined when it is assumed that the seeded rainfall $R_{s1}, \ldots, R_{sn}$ is a random sample of n observations from a gamma distribution with parameters α and β_s; and $R_{c1}, \ldots, R_{cm}$ is a random sample of m observations from a gamma distribution with parameters α and β_c. Note that the shape parameters are identical for the two populations.

The actual seeding factor is defined by $F = \langle R \rangle_s / \langle R \rangle_c = \beta_c/\beta_s$ and